View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of Waterloo's Institutional Repository

Preprocessing and Reduction for Semidefinite Programming via Facial Reduction: Theory and Practice

by

Yuen-Lam Cheung

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Combinatorics & Optimization

Waterloo, Ontario, Canada, 2013

© Yuen-Lam Cheung 2013

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Semidefinite programming is a powerful modeling tool for a wide range of optimization and feasibility problems. Its prevalent use in practice relies on the fact that a (nearly) optimal solution of a semidefinite program can be obtained efficiently in both theory and practice, provided that the semidefinite program and its dual satisfy the Slater condition.

This thesis focuses on the situation where the Slater condition (i.e., the existence of positive definite feasible solutions) does not hold for a given semidefinite program; the failure of the Slater condition often occurs in structured semidefinite programs derived from various applications. In this thesis, we study the use of the facial reduction technique, originally proposed as a theoretical procedure by Borwein and Wolkowicz, as a preprocessing technique for semidefinite programs. Facial reduction can be used either in an algorithmic or a theoretical sense, depending on whether the structure of the semidefinite program is known *a priori*.

The main contribution of this thesis is threefold. First, we study the numerical issues in the implementation of the facial reduction as an algorithm on semidefinite programs, and argue that each step of the facial reduction algorithm is backward stable. Second, we illustrate the theoretical importance of the facial reduction procedure in the topic of sensitivity analysis for semidefinite programs. Finally, we illustrate the use of facial reduction technique on several classes of structured semidefinite programs, in particular the side chain positioning problem in protein folding.

Acknowledgements

I would like to thank my advisor, Henry Wolkowicz, for all his guidance, assistance and patience. Henry has given me a lot of opportunities to explore different research directions and to attend various academic activities. The immense amount of time and energy that Henry has put in guiding me through our research projects has led to a very fruitful learning experience during my Ph.D.

I would like to thank my examination committee members, Forbes Burkowski, Levent Tunçel, Stephen Vavasis and Shuzhong Zhang for their useful feedback and comments on my thesis.

I would also like to thank the faculty and staff members of the Department of Combinatorics and Optimization as well as the Faculty of Mathematics. I have learned a lot over the years I have spent in University of Waterloo, thanks to the wonderful learning environment.

I appreciate the support from my family; without them, this accomplishment would not have been possible.

Table of Contents

List of Algorithms

List of Tables

Chapter 1

Introduction

Semidefinite programming, an important class of conic programming, is a powerful modelling tool that has a wide range of applications, see e.g., [5, 12, 22, 88, 90, 98]. In addition to its modelling power, one factor that contributes to the prevalent use of semidefinite programs is that they can be solved efficiently, both in theory and practice. Currently, the most popular algorithms for solving semidefinite programs are *interior point methods* [3, 65, 101], implemented by open-source software such as SeDuMi [85], SDPT3 [92], DSDP [10], CSDP [16, 15] and the SDPA family [42, 43, 100], as well as MOSEK, one of the first commercial software packages that include an SDP solver.

For the majority of interior point methods in the literature, the theoretical convergence results rely on the assumption that the semidefinite program (and its dual, if the interior point method uses both the primal and the dual) satisfies strict feasibility (or the *Slater condition*), which facilitates the well-definedness of the *central path*, an essential notion for the theoretical efficiency of interior point methods. (On the other hand, there are a small number of algorithms, such as [39], that provide convergence analysis without the strict feasibility assumption.)

In practice, it cannot be taken for granted that a given semidefinite program must be strictly feasible. One way to tackle the possible failure of strict feasibility is to reformulate the given semidefinite program into another semidefinite program, so that both the reformulated semidefinite program and its dual are *always* strictly feasible. Then the theory of interior point methods (such as the existence of the central path) remains relevant for the reformulated semidefinite program, even if the given semidefinite program is not strictly feasible (or even not feasible).

A possible reformulation is to add bound constraints on the variables [10]. Another popular reformulation is via the *homogeneous self-dual embedding* method, introduced in [57, 102] and

further studied in [30, 31, 63, 66, 101, 104]. The idea is to *embed* the given semidefinite program into a larger one by adding additional variables, so that the new larger semidefinite program and its dual are actually the same (hence the name self-dual); then an interior point method is applied on the new semidefinite program. Implementations of the self-dual embedding technique include SeDuMi [85] and [26].

In this thesis, we study an alternative way to preprocess semidefinite programs that are not strictly feasible, via *facial reduction*. The idea of facial reduction for semidefinite programming is that, the feasible region of a semidefinite program that is not strictly feasible (especially those arising from applications) must lie in a proper face of the cone of positive semidefinite matrices; therefore, we may restrict the semidefinite program to that proper face, allowing for a reduction in problem size due to the facial structure of the cone of positive semidefinite matrices. The resultant smaller equivalent semidefinite program can then be solved by standard solvers in a numerically stable manner.

The facial reduction technique was introduced in [18, 19, 20] for abstract convex programs, and has been specialized in linear conic programs [70, 86, 95] (see also [62, Section 7] for the dual version) as well as in SDP [27]. Ramana *et al.* [74] related the facial reduction on semidefinite programs with the *extended Lagrange-Slater dual*, which was introduced in [73]. Extensions and variants of the facial reduction algorithm exist for partially finite convex programs [61] and for doubly nonnegative relaxation of the mixed binary nonconvex quadratic optimization problems [87].

This thesis studies the use of the facial reduction technique as a preprocessing technique for semidefinite programs that fail the strict feasibility assumption. We are interested in the implementation of the facial reduction technique in both general semidefinite programs and special classes of structured semidefinite programs.

1.1 Is facial reduction necessary?

It has been proved that strict feasibility is a generic property for feasible linear conic programs:

Theorem 1.1.1. *[37, Theorem 3.2] For almost all problem instances of semidefinite programs in the form* (P) *(defined on Page 30), either one of the following holds:*

(1) (P) *is infeasible; or*

(2) (P) *is strictly feasible.*

In this light, one would think that the strict feasibility occurs frequently enough. Nonetheless, as we shall see shortly, a lot of structured semidefinite programming problems are indeed not strictly feasible, precisely because of their specific structures. In those cases, one can often regularize the semidefinite programs and obtain equivalent smaller problems that are strictly feasible. The solution of such smaller problems naturally takes less time, and are empirically more numerically stable (see, e.g., Section 9.5).

Even though the existing convergence theory for interior point methods typically depends on strict feasibility assumption, it does not mean that the software would have problems when solving a semidefinite program that is not strictly feasible. Nonetheless, in numerical tests it appears that standard software packages such as SDPT3 [92] and SeDuMi [85] do not always tackle semidefinite programs that are not strictly feasible (or *nearly so*) very well (particularly if the SDP instance has a positive duality gap), see e.g. [27, 49, 96]. In this light, we find it worthwhile to study the facial reduction technique as an alternative for handling the failure of strict feasibility.

1.2 Some applications

Semidefinite programming often arises from applications that are specially structured. In the applications listed below, the special structures of the semidefinite programs allow one to locate the *minimal face* of the semidefinite program *a priori*. Then the semidefinite program can be *facially reduced* to become a smaller and more stable problem (in the sense that strict feasibility is satisfied), prior to being solved by standard solvers.

Applications of facial reduction include:

- sensor network localization [60];
- preprocessing of semidefinite programming relaxation from
  - quadratic assignment problem (QAP) [105];
  - graph partitioning (GP) problem [99];
  - polynomial optimization: specifically, finding sparse sum-of-squares representations of polynomials [94];
- protein conformation [2, 24].

1.3 Facial reduction as a regularization technique and as an algorithm

By facial reduction, we mean either a theoretical regularization or a numerical algorithm.

- As a theoretical regularization technique, facial reduction means rewriting a semidefinite program using the *minimal face* of the cone of positive semidefinite matrices containing the feasible region, to get a smaller equivalent problem.

  This regularization procedure is often applicable to structured semidefinite programs arising from applications, and the minimal face often exposes some additional properties of the feasible region. For structured semidefinite programs, facial reduction can usually be done before the use of a standard solver for solving the semidefinite programs.

  An example is the semidefinite programming relaxation of the *side chain positioning problem*, a combinatorial optimization problem. (See Chapter 9.) Since the semidefinite program arising from any instance of the side chain positioning problem of fixed size always has the same feasible region, we can compute the minimal face of the semidefinite program, and make use of the minimal face to get an equivalent and regularized semidefinite program.

- As a numerical algorithm, facial reduction finds in each iteration a "smaller" face of the cone of positive semidefinite matrices that contains the feasible region of the given feasible semidefinite program, by solving an appropriate conic program (which is strictly feasible). The output of the algorithm is the data for a smaller equivalent semidefinite program that is strictly feasible.

  One may question the necessity of using a facial reduction algorithm to correct the failure of strict feasibility, when one can use, for instance, self-dual embedding [30, 31, 63, 104, 101] to ensure that any semidefinite program passed to a standard solver is strictly feasible. Empirically, SeDuMi often fails even on small scale semidefinite programs that are not strictly feasible and have positive duality gap, even though theoretically it should not be the case.

  The facial reduction algorithm aims at ensuring that a feasible semidefinite program can be solved in a numerically stable manner, even though it may take longer aggregate runtime to use the facial reduction algorithm together with the solution of the facially reduced SDP.

  (Naturally, we ask if the facial reduction algorithm does produce a numerically equivalent problem. We propose an implementation of the facial reduction as an algorithm in Chapter 5, and show in Chapter 6 that each iteration is backward stable.)

1.4 Contribution of the thesis

This thesis explores three different aspects of facial reduction.

(1) In Chapters 5 and 6, we propose an implementation of facial reduction as an algorithm. We study the numerical issues of the implementation, and argue that each iteration of the facial reduction algorithm proposed is backward stable.

(2) In Chapter 7, we give an example highlighting the theoretical importance of the facial reduction algorithm. It has been shown [86] that the number of facial reduction iterations plays an important role in the error bound for linear matrix inequalities. We use this result to prove a perturbation result on semidefinite programs, that the change in the optimal value due to a feasible perturbation of the right-hand side of the constraints in a semidefinite program may also depend on the number of facial reduction iterations.

(3) In Chapters 8 and 9, we illustrate the use of the facial reduction technique on several classes of structured semidefinite programs, in particular the side chain positioning problem in protein folding. We obtain a stable semidefinite programming relaxation of the side chain positioning problem, that can be solved efficiently in numerical tests.

1.5 Organization of the thesis

This thesis has three parts. Part I collects standard results in the literature. Preliminaries on convex analysis (in Chapter 2) and conic programming (in Chapter 3), as well as the basics on facial reduction for linear conic programs (in Chapter 4) shall be able to prepare a reader unfamiliar with the topics for the following chapters.

In Part II, we study the implementation of a facial reduction algorithm for semidefinite programs. Chapter 5 focuses on the implementation details and the numerical issues, as well as reports some numerical results on the implementation of the facial reduction algorithm in comparison with the solution of semidefinite programs by SeDuMi without preprocessing. (We find that, empirically, numerical instability occasionally occurs when solving semidefinite programs that are not strictly feasible with SeDuMi.) In Chapter 6, we prove that one iteration of the facial reduction algorithm is backward stable.

In Part III, we demonstrate some uses of the facial reduction technique. Chapter 7 studies the use of facial reduction to provide sensitivity analysis on semidefinite programs. This result relies

on an earlier result concerning the error bound for linear matrix inequalities proved by Sturm [86], which also made use of facial reduction. Chapter 8 gives an overview of classes of problems that are not strictly feasible and can be regularized by facial reduction. Chapter 9 studies in detail a class of such problems, the side chain positioning problem, and illustrates how one can find the minimal face of the feasible region of the semidefinite program. Numerics show that the use of the facial reduction improves both the runtime and the solution quality.

Part I

Preliminaries

Chapter 2

Preliminaries on convex analysis

This chapter provides a summary of relevant notions and results from convex analysis. These can be found in standard textbooks such as [54] and [79].

Let $\mathbb{R}$ denote the set of all real numbers. Let $\mathbb{V}$ be a vector space over the reals. An *inner product* on $\mathbb{V}$ is a map $\langle \cdot, \cdot \rangle_{\mathbb{V}} : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ that satisfies the following conditions:

(1) $\langle x, x \rangle_{\mathbb{V}} \geq 0$ for all $x \in \mathbb{V}$, and equality holds if and only if $x = 0$;

(2) $\langle x, y \rangle_{\mathbb{V}} = \langle y, x \rangle_{\mathbb{V}}$ for all $x, y \in \mathbb{V}$; and

(3) $\langle \alpha x + \beta y, z \rangle_{\mathbb{V}} = \alpha \langle x, z \rangle_{\mathbb{V}} + \beta \langle y, z \rangle_{\mathbb{V}}$ for all $\alpha, \beta \in \mathbb{R}$ and $x, y, z \in \mathbb{V}$.

A vector space $\mathbb{V}$ endowed with an inner product $\langle \cdot, \cdot \rangle_{\mathbb{V}}$ is called an *inner product space*, denoted by $(\mathbb{V}, \langle \cdot, \cdot \rangle_{\mathbb{V}})$. Define $\|x\|_{\mathbb{V}} := \sqrt{\langle x, x \rangle_{\mathbb{V}}}$ for all $x \in \mathbb{V}$. The map $\| \cdot \|_{\mathbb{V}} : \mathbb{V} \to \mathbb{R}$ is a *norm*, i.e., $\| \cdot \|_{\mathbb{V}}$ satisfies the following criteria:

(1) $\|x\|_{\mathbb{V}} \geq 0$ for all $x \in \mathbb{V}$ and equality holds if and only if $x = 0$;

(2) $\|\alpha x\|_{\mathbb{V}} = |\alpha| \|x\|_{\mathbb{V}}$ for all $\alpha \in \mathbb{R}$; and

(3) $\|x + y\|_{\mathbb{V}} \leq \|x\|_{\mathbb{V}} + \|y\|_{\mathbb{V}}$ for all $x, y \in \mathbb{V}$. The last criterion is called the *triangle inequality*.

If $(\mathbb{V}_1, \langle \cdot, \cdot \rangle_{\mathbb{V}_1}), (\mathbb{V}_2, \langle \cdot, \cdot \rangle_{\mathbb{V}_2})$ are inner product spaces, then the *direct product* of $\mathbb{V}_1$ and $\mathbb{V}_2$, i.e., the vector space defined by $\mathbb{V}_1 \times \mathbb{V}_2 := \left\{ (x^{(1)}, x^{(2)}) : x^{(1)} \in \mathbb{V}_1, x^{(2)} \in \mathbb{V}_2 \right\}$ with

$$(x^{(1)}, x^{(2)}) + \lambda (y^{(1)}, y^{(2)}) := (x^{(1)} + \lambda y^{(1)}, x^{(2)} + \lambda y^{(2)})$$

for all $(x^{(1)}, x^{(2)}), (y^{(1)}, y^{(2)}) \in \mathbb{V}_1 \times \mathbb{V}_2$ and scalar λ, can be equipped with the standard inner product $\langle (x^{(1)}, x^{(2)}), (y^{(1)}, y^{(2)}) \rangle_{\mathbb{V}_1 \times \mathbb{V}_2} := \langle x^{(1)}, y^{(1)} \rangle_{\mathbb{V}_1} + \langle x^{(2)}, y^{(2)} \rangle_{\mathbb{V}_2}$. The *Cartesian product* of two nonempty sets $\mathcal{S}_1 \subseteq \mathbb{V}_1$ and $\mathcal{S}_2 \subseteq \mathbb{V}_2$ is defined as the set

$$\mathcal{S}_1 \times \mathcal{S}_2 := \left\{ (x^{(1)}, x^{(2)}) : x^{(1)} \in \mathcal{S}_1, x^{(2)} \in \mathcal{S}_2 \right\} \subseteq \mathbb{V}_1 \times \mathbb{V}_2.$$

For any $x \in \mathbb{V}$ and $\delta > 0$, define $B(x, \delta) := \{y \in \mathbb{V} : \|y - x\|_{\mathbb{V}} \leq \delta\}$. A set $\mathcal{S} \subseteq \mathbb{V}$ is said to be *open* if for all $x_0 \in \mathbb{V}$, there exists $\delta > 0$ such that $B(x_0, \delta) \subseteq \mathcal{S}$. A set $\mathcal{S}$ is said to be *closed* if its complement $\mathbb{V} \backslash \mathcal{S}$ is open. The *interior* of a set $\mathcal{S} \subseteq \mathbb{V}$ is the (inclusion-wise) maximal open subset of $\mathcal{S}$, denoted by $\text{int}(\mathcal{S})$.

The *Minkowski sum* of two nonempty sets $\mathcal{S}_1$ and $\mathcal{S}_2$ is $\mathcal{S}_1 + \mathcal{S}_2 := \{s_1 + s_2 : s_i \in \mathcal{S}_i,\ i = 1, 2\}$, and $\alpha \mathcal{S}_1 := \{\alpha s : s \in \mathcal{S}_1\}$ for $\alpha \in \mathbb{R}$. When $\mathcal{S}_1 = \{s\}$ is a singleton, we write $s + \mathcal{S}_2 := \mathcal{S}_1 + \mathcal{S}_2$.

For any nonempty set $\mathcal{Y} \subseteq \mathbb{V}$ and $Y \in \mathbb{V}$, define

$$\text{dist}(Y, \mathcal{Y}) := \inf_X \{\|Y - X\|_{\mathbb{V}} : X \in \mathcal{Y}\}.$$

2.1 Convex cones

A set $\mathcal{K} \subseteq \mathbb{V}$ is called a *linear subspace* of $\mathbb{V}$ if $\alpha x + \beta y \in \mathcal{K}$ for all $x, y \in \mathcal{K}$ and $\alpha, \beta \in \mathbb{R}$. A set $\mathcal{K} \subseteq \mathbb{V}$ is said to be *affine* if $\alpha x + (1 - \alpha) y \in \mathcal{K}$ for all $x, y \in \mathcal{K}$ and $\alpha \in \mathbb{R}$, *convex* if $\alpha x + (1 - \alpha) y \in \mathcal{K}$ for all $x, y \in \mathcal{K}$ and $\alpha \in [0, 1]$, and a *cone* if $\alpha x \in \mathcal{K}$ for all $x \in \mathcal{K}$ and $\alpha > 0$. The *affine hull* and *convex hull* of a nonempty set $\mathcal{S} \subseteq \mathbb{V}$ are respectively defined as

$$\text{aff}(\mathcal{S}) := \{\beta x + (1 - \beta) y : x, y \in \mathcal{S},\ \beta \in \mathbb{R}\},$$
$$\text{conv}(\mathcal{S}) := \{\beta x + (1 - \beta) y : x, y \in \mathcal{S},\ \beta \in [0, 1]\}.$$

A set $\mathcal{K} \subseteq \mathbb{V}$ is an affine set if and only if $\mathcal{K} = x + \mathcal{L}$ for some $x \in \mathcal{K}$ and a linear subspace $\mathcal{L} \subseteq \mathbb{V}$. The *dimension* of an affine set $\mathcal{K} = x + \mathcal{L}$, where $\mathcal{L}$ is a linear subspace, is defined as the dimension of the linear subspace $\mathcal{L}$. The *dimension* of a nonempty set $\mathcal{S}$ is the dimension of the affine hull $\text{aff}(\mathcal{S})$. Several commonly used objects are defined based on their specified dimension. For instance, an affine set in $\mathbb{V}$ with dimension $\dim(\mathbb{V}) - 1$ is called a *hyperplane*, and a cone of dimension one is called a *ray*. The *relative interior* of a nonempty convex set $\mathcal{C}$ is the set

$$\text{ri}(\mathcal{C}) := \{x \in \mathcal{C} : \exists\, \epsilon > 0\ B(x, \epsilon) \cap \text{aff}(\mathcal{C}) \subseteq \mathcal{C}\}.$$

A nonempty convex set $\mathcal{C}$ may have empty interior but $\mathrm{ri}(\mathcal{C})$ is always nonempty. If $\mathrm{int}(\mathcal{C}) \neq \varnothing$, then $\mathrm{ri}(\mathcal{C}) = \mathrm{int}(\mathcal{C})$. We say that a set $\mathcal{C}$ is *relatively open* if $\mathrm{ri}(\mathcal{C}) = \mathcal{C}$. If $\mathcal{C}$ is convex (resp. conic), then $\mathrm{ri}(\mathcal{C})$ is convex (resp. conic) too.[1]

$\mathcal{K}$ is said to be a *proper cone* if $\mathcal{K}$ is a closed convex cone with nonempty interior and $\mathcal{K}$ is *pointed*, i.e., the *lineality space* of $\mathcal{K}$, which is defined as $\mathcal{K} \cap (-\mathcal{K})$, equals $\{0\}$. A proper cone $\mathcal{K} \subseteq \mathbb{V}$ can induce a *partial ordering* $\succeq_{\mathcal{K}}$ on $\mathbb{V}$:

$$\begin{aligned} x \succeq_{\mathcal{K}} y \quad &\Longleftrightarrow \quad x - y \in \mathcal{K}, \\ x \succ_{\mathcal{K}} y \quad &\Longleftrightarrow \quad x - y \in \mathrm{int}(\mathcal{K}). \end{aligned}$$

It is easy to check that $\succeq_K$ is indeed a partial order:

- for any $x \in \mathbb{V}$, $x - x = 0 \in \mathcal{K}$, so $x \succeq_{\mathcal{K}} x$;
- for any $x, y \in \mathbb{V}$, if $x \succeq_{\mathcal{K}} y$ and $y \succeq_{\mathcal{K}} x$, then $x - y \in \mathcal{K} \cap (-\mathcal{K}) = \{0\}$, so $x = y$;
- for any $x, y, z \in \mathbb{V}$ with $x \succeq_{\mathcal{K}} y$ and $y \succeq_{\mathcal{K}} z$, we have $x - z = (x - y) + (y - z) \in \mathcal{K}$, so $x \succeq_{\mathcal{K}} z$.

Moreover, when equipped with the partial order $\succeq_K$, $\mathbb{V} = (\mathbb{V}, \succeq_K)$ becomes an *ordered vector space*:

- for any $x, y, z \in \mathbb{V}$ with $x \succeq_{\mathcal{K}} y$, we have $(x - z) - (y - z) = x - y \in \mathcal{K}$, so $x - z \succeq_{\mathcal{K}} y - z$;
- for any $x, y \in \mathbb{V}$ with $x \succeq_{\mathcal{K}} y$, if $0 \leq \alpha \in \mathbb{R}$, then $\alpha(x - y) \in \mathcal{K}$, so $\alpha x \succeq_{\mathcal{K}} \alpha y$.

The *dual cone* of any set $\mathcal{S}$ (with respect to inner product $\langle \cdot, \cdot \rangle_{\mathbb{V}}$) is defined as

$$\mathcal{S}^* := \{y \in \mathbb{V} : \langle y, x \rangle_{\mathbb{V}} \geq 0, \ \forall\, x \in \mathcal{S}\}.$$

A dual cone is closed, convex and conic (even if $\mathcal{S}$ is not). $\mathcal{K}$ is said to be *self-dual* if $\mathcal{K}^* = \mathcal{K}$.

Conic programs often require the variables to lie in a Cartesian product of multiple convex cones; for this reason we point out a few basic facts about the Cartesian product of sets. For any pair of nonempty sets $\mathcal{S}_1 \subseteq \mathbb{V}_1$ and $\mathcal{S}_2 \subseteq \mathbb{V}_2$, the dual cone of $\mathcal{S}_1 \times \mathcal{S}_2$ is $(\mathcal{S}_1 \times \mathcal{S}_2)^* = \mathcal{S}_1^* \times \mathcal{S}_2^*$. If $\mathcal{S}_1$ and $\mathcal{S}_2$ are affine (resp. convex or conic), then their Cartesian product $\mathcal{S}_1 \times \mathcal{S}_2 \subseteq \mathbb{V}_1 \times \mathbb{V}_2$ is

[1] We only prove the conic case. Let $\mathcal{F}$ be a nonempty conic set, and let $Z \in \mathrm{ri}(\mathcal{F})$. If $\alpha \in (0, 1)$, then $\beta := \frac{1-\alpha}{1-\alpha/2} \in (0, 1)$ and $\alpha Z = \beta\left(\frac{1}{2}\alpha Z\right) + (1 - \beta)Z$. But $Z \in \mathrm{ri}(\mathcal{F})$ and $\frac{1}{2}\alpha Z \in \mathcal{F}$. So $\alpha Z \in \mathrm{ri}(\mathcal{F})$.

If $\alpha > 1$, then $\beta := \frac{\alpha-1}{2\alpha-1} \in (0, 1)$ and $\alpha Z = \beta(2\alpha Z) + (1 - \beta)Z$. But $Z \in \mathrm{ri}(\mathcal{F})$ and $2\alpha Z \in \mathcal{F}$. So $\alpha Z \in \mathrm{ri}(\mathcal{F})$.

also affine (resp. convex or conic). If $\mathcal{S}_1$ and $\mathcal{S}_2$ are proper cones, then $\mathcal{S}_1 \times \mathcal{S}_2$ is a proper cone too.

We state a very important result in convex analysis, that two "disjoint" convex sets can be separated by a hyperplane.

Theorem 2.1.1 (Separation theorem, version 1; Theorem 11.2, [79])**.** *Let $\mathcal{C} \subseteq \mathbb{V}$ be a nonempty convex set such that $\mathrm{ri}(\mathcal{C}) = \mathcal{C}$. Let $\mathcal{M} \subseteq \mathbb{V}$ be an affine set. If $\mathcal{M} \cap \mathcal{C} = \varnothing$, then there exist $0 \neq z \in \mathbb{V}$ and $\beta \in \mathbb{R}$ such that*

$$\langle z, x \rangle = \beta, \ \forall\, x \in \mathcal{M}; \ \langle z, y \rangle > \beta, \ \forall\, y \in \mathcal{C}.$$

Theorem 2.1.2 (Separation theorem, version 2; Theorem 11.3, [79])**.** *Let $\mathcal{C}_1, \mathcal{C}_2 \subseteq \mathbb{V}$ be nonempty convex sets. If $\mathrm{ri}(\mathcal{C}_1) \cap \mathrm{ri}(\mathcal{C}_2) = \varnothing$, then there exist $0 \neq z \in \mathbb{V}$ and $\beta \in \mathbb{R}$ such that*

$$\langle z, x \rangle \geq \beta \geq \langle z, y \rangle \ \forall\, x \in \mathcal{C}_1, \ y \in \mathcal{C}_2,$$

and $\sup_{x \in \mathcal{C}_1} \langle z, x \rangle > \beta$. (The converse is also true.)

One application of the separation theorem is to establish the bipolar theorem:

Theorem 2.1.3. *For any nonempty set $\mathcal{K} \subseteq \mathbb{V}$, $\mathcal{K}$ is a closed convex cone if and only if $\mathcal{K}^{**} = \mathcal{K}$.*

Using Theorem 2.1.3, we can show that the dual cone of a proper cone is also a proper cone.

Proposition 2.1.4. *Let $\mathcal{K} \subset \mathbb{V}$ be a proper cone. Then $\mathcal{K}^*$ is also a proper cone.*

Proof. $\mathcal{K}$ being a convex cone implies that $\mathcal{K}^*$ is a closed convex cone. We need to show that $\mathcal{K}^*$ has nonempty interior and is pointed. If $\mathcal{K}^*$ has empty interior, then the linear subspace $\mathrm{aff}(\mathcal{K}^*) \neq \mathbb{V}$. Hence there exist a nonzero $x \in (\mathrm{aff}(\mathcal{K}^*))^\perp \subseteq (\mathcal{K}^*)^\perp \subseteq \mathcal{K}^{**} \cap (-\mathcal{K}^{**}) = \mathcal{K} \cap (-\mathcal{K})$ by Theorem 2.1.3. But $\mathcal{K}$ is pointed so x must be zero. This contradiction implies that $\mathcal{K}^*$ cannot have empty interior. To see that $\mathcal{K}^*$ is pointed, pick any $x \in \mathcal{K}^* \cap (-\mathcal{K}^*)$. Then $\langle x, y \rangle_{\mathbb{V}} = 0$ for all $y \in \mathcal{K}$, i.e., $x \in \mathcal{K}^\perp$. But $\mathcal{K}$ has nonempty interior, so x must be zero. Hence $\mathcal{K}^* \cap (-\mathcal{K}^*) = \{0\}$, i.e., $\mathcal{K}$ is pointed. □

2.1.1 Examples

In this section we review a few well-known examples of finite dimensional inner product spaces and closed convex cones in these spaces. First we introduce some notations.

Notation

We adopt the MATLAB notation $j:k := \{j, j+1, \ldots, k\}$ for integers $j < k$ (so $i \in j:k$ means $i \in \{j, j+1, \ldots, k\}$). Let $\mathbb{R}^n$ be the set of all real n-vectors, and $\mathbb{R}^{m\times n}$ be the set of all real $m \times n$ matrices. We denote the identity matrix in $\mathbb{R}^{n\times n}$ by I and the matrix of all ones in $\mathbb{R}^{m\times n}$ by J. The *transpose* of a square matrix $X = [X_{ij}]_{i,j=1:n} \in \mathbb{R}^{n\times n}$ is defined as $X^\top := [X_{ji}]_{i,j=1:n}$. We denote the j-th column of a matrix X by $X_{:j}$ and similarly the i-th row of X by $X_{i:}$. We say that a matrix $X \in \mathbb{R}^{m\times n}$ has *orthonormal columns* if $(X_{:j})^\top X_{:j} = 1$ and $(X_{:i})^\top X_{:j} = 0$ for all $i \neq j \in 1:n$, or equivalently, $X^\top X = I$. A real square matrix X is said to be an *orthogonal matrix* if $X^\top X = I = XX^\top$. We define the *trace* of a square matrix X to be the sum $\sum_j X_{jj}$ of the diagonal entries.

Let $\mathbb{S}^n \subset \mathbb{R}^{n\times n}$ denote the set of all symmetric matrices, i.e., matrices X satisfying $X = X^\top$. Then $\mathbb{S}^n$ is a vector subspace of $\mathbb{R}^{n\times n}$. We will often make use of the linear map

$$\operatorname{diag} : \mathbb{S}^n \to \mathbb{R}^n : [X_{ij}]_{i,j=1:n} \mapsto (X_{11}, X_{22}, \ldots, X_{nn})$$

and its adjoint $\operatorname{Diag} := \operatorname{diag}^*$, which takes a vector $x \in \mathbb{R}^n$ and forms a diagonal matrix with x as the diagonal.

We will use $\bar{e}$ to denote the vector of all ones of appropriate length, and $e_j \in \mathbb{R}^{n\times n}$ to denote the j-th standard unit vector, i.e., e_j is the j-th column of the $n \times n$ identity matrix.

Nonnegative orthant and second-order cone

In $\mathbb{R}^n$ (equipped with the usual inner product $x^\top y := \sum_k x_k y_k$ and the induced Euclidean norm $\|x\| := \sqrt{x^\top x}$), the most common example of a closed convex cone is the *nonnegative orthant* $\mathbb{R}^n_+ := \{x \in \mathbb{R}^n : x_i \geq 0,\ \forall\, i = 1:n\}$.

If $n \geq 2$, another common example is the *second-order cone*:

$$\mathcal{Q}^n := \left\{ \begin{pmatrix} \alpha \\ z \end{pmatrix} \in \mathbb{R}^n : \|z\| \leq \alpha,\ \alpha \in \mathbb{R},\ z \in \mathbb{R}^{n-1} \right\}.$$

Both $\mathbb{R}^n_+$ and $\mathcal{Q}^n$ are proper and self-dual cones.

Positive semidefinite cone and copositive cone

We introduce the vector space $\mathbb{S}^n$ of $n \times n$ real symmetric matrices. In $\mathbb{S}^n$ we can define the positive semidefinite cone and the copositive cone. The positive semidefinite cone will play a central role in this thesis.

A square matrix $X \in \mathbb{R}^{n\times n}$ is said to be *symmetric* if $X = X^\top$. A real symmetric matrix X is always *orthogonally diagonalizable*: there exist an orthogonal matrix $U \in \mathbb{R}^{n\times n}$ and a diagonal matrix $D \in \mathbb{R}^{n\times n}$ such that $X = UDU^\top$. The factorization $UDU^\top$ is called the *spectral decomposition* of X. The diagonal entries $D_{11}, \ldots, D_{nn}$ are called the *eigenvalues* of X, and the vector $U_{:j}$ is called the *eigenvector* corresponding to eigenvalue D_{jj}, for $j \in 1:n$. Assuming $D_{11} \geq D_{22} \geq \cdots \geq D_{nn}$, the function $\lambda : \mathbb{S}^n \to \mathbb{R}^n : X \mapsto (D_{11}, D_{22}, \ldots, D_{nn})$ is well defined. In addition, the largest eigenvalue function $\lambda_{\max}(X) := \max_j \{D_{jj} : j \in 1:n\}$ and the smallest eigenvalue function $\lambda_{\min}(X) := \min_j \{D_{jj} : j \in 1:n\}$ are also well-defined. A real symmetric matrix whose eigenvalues are all nonnegative (resp., all positive) is called a *positive semidefinite* matrix (resp., a *positive definite* matrix). There are many different characterizations of positive semidefinite matrices (see e.g. [90]). If $X = UDU^\top$ is the spectral decomposition of a positive semidefinite matrix X and $D_{\bar{n}\bar{n}} > 0 = D_{jj}$ for $j \in (\bar{n}+1):n$, then $X = UDU^\top = PD_+P^\top$, where

$$D = \begin{array}{c} \\ \bar{n} \\ n-\bar{n} \end{array} \begin{array}{c} \begin{array}{cc} \bar{n} & n-\bar{n} \end{array} \\ \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} \end{array}, \quad D_+ = \begin{bmatrix} D_{11} & 0 & \cdots & 0 \\ 0 & D_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & D_{\bar{n}\bar{n}} \end{bmatrix}, \quad U = \begin{array}{c} \\ n \end{array} \begin{array}{c} \begin{array}{cc} \bar{n} & n-\bar{n} \end{array} \\ \begin{bmatrix} P & Q \end{bmatrix} \end{array}.$$

We call the factorization $PD_+P^\top$ the *compact spectral decomposition* of X.

We mention a frequently used result for checking the positive definiteness (or semidefiniteness) of a matrix, using the *Schur complement*.

Theorem 2.1.5. *Let* $M = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \in \mathbb{S}^{s+t}$. *If* $A \succ 0$, *then the Schur complement of* M *with respect to the (1,1)-block* A, *defined as the matrix* $C - BA^{-1}B^\top$, *is positive semidefinite if and only if* M *is positive semidefinite.*

Theorem 2.1.5 can be phrased in a more general form, see, e.g.,[22, 103]. More details about positive semidefinite matrices can be found in, e.g., [22, 90] and [98, Chapter 2].

The set $\mathbb{S}^n$ of all real symmetric matrices can be endowed with an inner product. For any $X, Y \in \mathbb{S}^n$, we define

$$\langle X, Y\rangle := \operatorname{tr}(XY) = \sum_{i,j=1}^{n} X_{ij}Y_{ij}.$$

It is easy to check that $\langle\cdot,\cdot\rangle$ is indeed an inner product: for any $X \in \mathbb{S}^n$, $\langle X, X\rangle = \sum_{i,j=1}^n (X_{ij})^2$ is obviously nonnegative, and equals zero if and only if $X = 0$. Moreover, for any $X, Y, Z \in \mathbb{S}^n$ and

$\alpha, \beta \in \mathbb{R}$, we have $\langle X, Y\rangle = \langle Y, X\rangle$ and $\langle X, \alpha Y + \beta Z\rangle = \alpha\langle X, Y\rangle + \beta\langle X, Z\rangle$. Therefore, $(\mathbb{S}^n, \langle\cdot,\cdot\rangle)$ is an inner product space. We denote the induced norm on $\mathbb{S}^n$ by $\|X\| := \sqrt{\langle X, X\rangle}$.

Having endowed $\mathbb{S}^n$ with an inner product, we consider some examples of closed convex cones in $\mathbb{S}^n$. The set of all positive semidefinite matrices forms a convex cone in $\mathbb{S}^n$, and we call this set the *positive semidefinite cone*, denoted by $\mathbb{S}^n_+$. In addition, we denote by $\mathbb{S}^n_{++}$ the set of all positive definite matrices, and $\mathbb{S}^n_{++}$ also forms a convex cone in $\mathbb{S}^n$. For any $X, Y \in \mathbb{S}^n$, we write $X \succeq Y$ to mean that $X - Y \in \mathbb{S}^n_+$ and $X \succ Y$ to mean that $X - Y \in \mathbb{S}^n_{++}$ It is well-known that $\mathbb{S}^n_+$ is a closed convex cone in $\mathbb{S}^n$ and $\operatorname{int}(\mathbb{S}^n_+) = \mathbb{S}^n_{++}$. Moreover, $\mathbb{S}^n_+$ is self-dual.

The *copositive cone* $\mathcal{C}^n \subset \mathbb{S}^n$ is defined as the set of all matrices C that satisfy

$$x^\top C x \geq 0, \ \forall\, x \in \mathbb{R}^n_+.$$

Observe that $\mathbb{S}^n_+ \subset \mathcal{C}^n$ and $\mathcal{C}^n$ is a closed convex cone. $\mathcal{C}^n$ is not self-dual; indeed,

$$(\mathcal{C}^n)^* = \left\{ X \in \mathbb{S}^n : X = \sum_{k=1}^{m} \lambda_k q_k q_k^\top \text{ for some } \lambda \in \mathbb{R}^m_+,\ q_1, q_2, \ldots, q_m \in \mathbb{R}^n_+,\ m \geq 0 \text{ integer} \right\}$$

and is called the *completely positive cone*. Dickinson's thesis [33] provides a very comprehensive review on the known properties of copositive cones and completely positive cones.

2.2 Faces of a convex set

In this section we study an important notion concerning convex cones: their faces. The importance lies in the fact that they describe a hierarchy of boundary objects relevant in conic programs. Failure of commonly used constraint qualifications is often equivalent to the feasible region being contained in a proper face of the convex cone at hand, and knowledge about faces of that cone is essential for dealing with such situations.

We first define a face of a convex set.

Definition 2.2.1. *Let $\mathcal{S}$ be a nonempty convex set. A nonempty convex subset $\mathcal{F} \subseteq \mathcal{S}$ is said to be a* face *of $\mathcal{S}$ if*

$$x, y \in \mathcal{S} \quad \textit{and} \quad \alpha x + (1-\alpha) y \in \mathcal{F} \textit{ for some } 0 < \alpha < 1 \implies x, y \in \mathcal{F}.$$

We write $\mathcal{F} \trianglelefteq \mathcal{S}$ if $\mathcal{F}$ is a face of $\mathcal{S}$. A set $\mathcal{F}$ is said to be a proper face *of $\mathcal{S}$ if $\mathcal{F} \neq \mathcal{S}$ and $\mathcal{F} \trianglelefteq \mathcal{S}$, and is denoted by $\mathcal{F} \triangleleft \mathcal{S}$. If $z \in \mathcal{S}$ satisfies $\{z\} \trianglelefteq \mathcal{S}$, then z is said to be an* extreme point *of $\mathcal{S}$. If $\mathcal{F} \trianglelefteq \mathcal{S}$ and $\dim(\mathcal{F}) = \dim(\mathcal{S}) - 1$, then $\mathcal{F}$ is called a* facet *of $\mathcal{S}$. If*

$$\mathcal{F} = \mathcal{S} \cap \{x \in \mathbb{V} : \langle a, x\rangle_{\mathbb{V}} = \alpha\} \quad \textit{and} \quad \mathcal{S} \subseteq \{x \in \mathbb{V} : \langle a, x\rangle_{\mathbb{V}} \geq \alpha\}$$

for some $0 \neq a \in \mathbb{V}$ *and* $\alpha \in \mathbb{R}$*, then* $\mathcal{F} \trianglelefteq \mathcal{S}$ *and is said to be an* exposed face*.*

In particular, if $\mathcal{S}$ *is a convex cone, then any one-dimensional face of* $\mathcal{S}$ *is called an* extreme ray*. If an extreme ray of* $\mathcal{S}$ *is also an exposed face, then we call that face an* exposed ray*.*

Remark. In many texts (e.g. [12, 79]), the definition of a face allows it to be an empty set. We adopt the definition from [54, Definition 2.3.6], that a face must be a nonempty set.

We give some basic properties of the faces of convex cones and convex sets.

Proposition 2.2.2. *Let* $\mathcal{K}$ *be a convex cone containing 0 and* $\mathcal{S}$ *be a nonempty convex set. Then the following holds:*

1. $\mathcal{K}$ *is a face of itself, and any face* $\mathcal{F}$ *of* $\mathcal{K}$ *contains 0. Moreover, if* $\mathcal{K}$ *is pointed, then* $\{0\}$ *is a face of* $\mathcal{K}$*.*

2. *Let* $\varnothing \neq \mathcal{F} \subseteq \mathcal{K}$ *be a convex set.* $\mathcal{F} \subseteq \mathcal{K}$ *is a face of* $\mathcal{K}$ *if and only if*

$$s \in \mathcal{F} \text{ and } 0 \preceq_{\mathcal{K}} u \preceq_{\mathcal{K}} s \implies \beta u \in \mathcal{F} \ \forall\, \beta \geq 0. \tag{2.1}$$

   In particular,

$$\mathcal{F} \trianglelefteq \mathcal{K} \implies \mathcal{F} \text{ is a convex cone.}$$

3. *Let* $\varnothing \neq \mathcal{F} \subseteq \mathcal{K}$ *be a convex set.* $\mathcal{F} \subseteq \mathcal{K}$ *is a face of* $\mathcal{K}$ *if and only if*

$$x, y \in \mathcal{K} \quad \text{and} \quad x + y \in \mathcal{F} \implies x, y \in \mathcal{F}. \tag{2.2}$$

4. *If* $\mathcal{F}_1, \mathcal{F}_2$ *are faces of* $\mathcal{S}$*, then* $\mathcal{F}_1 \cap \mathcal{F}_2$ *is also a face of* $\mathcal{S}$*.*

5. *If* $\mathcal{F}_i$*,* $\mathcal{G}_i$ *are faces of* $\mathcal{S}$ *and* $\mathcal{F}_i \trianglelefteq \mathcal{G}_i \trianglelefteq \mathcal{S}$ *for* $i = 1, 2$*, then* $\mathcal{F}_1 \cap \mathcal{F}_2 \trianglelefteq \mathcal{G}_1 \cap \mathcal{G}_2$*.*

6. *If* $\mathcal{F}_1 \trianglelefteq \mathcal{F}_2$ *and* $\mathcal{F}_2 \trianglelefteq \mathcal{S}$*, then* $\mathcal{F}_1 \trianglelefteq \mathcal{S}$*.*

7. *If* $\mathcal{F}_1 \subseteq \mathcal{F}_2$ *are nonempty convex subsets of* $\mathcal{S}$ *and* $\mathcal{F}_1 \trianglelefteq \mathcal{S}$*, then* $\mathcal{F}_1 \trianglelefteq \mathcal{F}_2$*.*

8. $\mathcal{F}$ *is an exposed face of* $\mathcal{K}$ *if and only if* $\mathcal{F} = \mathcal{K} \cap \{a\}^{\perp}$ *for some* $0 \neq a \in \mathcal{K}^*$*.*

Proof. 1. That $\mathcal{K}$ is a face of itself is immediate from the definition of faces. To see that $\mathcal{F} \trianglelefteq \mathcal{K}$ implies $0 \in \mathcal{F}$, we suppose without loss of generality that $\mathcal{F} \neq \{0\}$. Let $0 \neq x \in \mathcal{F}$. Then $x = \frac{1}{2}(2x + 0)$ and $0, 2x \in \mathcal{K}$. Hence $0 \in \mathcal{F}$ by definition of a face.

Suppose that $\mathcal{K}$ is pointed. If $x, y \in \mathcal{K}$ satisfy $\frac{1}{2}(x + y) = 0$, then $x = -y \in -\mathcal{K}$. Hence $x \in \mathcal{K} \cap (-\mathcal{K}) = \{0\}$, i.e., $x = 0$.

2. Assume that $\mathcal{F} \trianglelefteq \mathcal{K}$. Let $s \in \mathcal{F}$ and u, $s-u \in \mathcal{K}$. Since $\mathcal{K}$ is a convex cone, observe that $s-\gamma u = s-u+(1-\gamma)u \in \mathcal{K}$ for all $\gamma \in [0,1]$. Now we show that $\beta u \in \mathcal{F}$ for all $\beta > 0$. Let $\alpha = \min\left\{\frac{1}{2}, \frac{1}{\beta}\right\} \in (0,1)$. Then $\alpha\beta \leq 1$, so $\frac{1}{1-\alpha}(s-\alpha\beta u) \in \mathcal{K}$. Hence $\alpha(\beta u) + (1-\alpha)\left(\frac{1}{1-\alpha}(s-\alpha\beta u)\right) = s \in \mathcal{F}$, implying that $\beta u \in \mathcal{F}$.

   Conversely, suppose that (2.1) holds. Let $x, y \in \mathcal{K}$ and $\alpha \in (0,1)$ satisfy $s := \alpha x + (1-\alpha)y \in \mathcal{F}$. Note that $s-(1-\alpha)y = \alpha x \in \mathcal{K}$ and $(1-\alpha)y \in \mathcal{K}$, so $y \in \mathcal{F}$ by (2.1). Similarly, $x \in \mathcal{F}$ too.

   Finally, to see that any face $\mathcal{F}$ has to be a cone, fix any $u \in \mathcal{F}$. Then $0 \leq_{\mathcal{K}} \frac{1}{2}u \leq_{\mathcal{K}} u$, and by (2.1) we have $\beta\left(\frac{1}{2}u\right) \in \mathcal{F}$ for all $\beta \geq 0$. Therefore $\beta u \in \mathcal{F}$ for all $\beta \geq 0$.

3. Assume that $\mathcal{F} \trianglelefteq \mathcal{K}$. Let $x, y \in \mathcal{K}$. Then $2x, 2y \in \mathcal{K}$ too. If $\mathcal{F} \ni x+y = \frac{1}{2}(2x+2y)$, then $2x, 2y \in \mathcal{F}$ too. But $\mathcal{F}$ is a cone, so $x, y \in \mathcal{F}$.

   Conversely, suppose that (2.2) holds. Let $s \in \mathcal{F}$ and $0 \preceq_{\mathcal{K}} u \preceq_{\mathcal{K}} s$. Then $u, s-u \in \mathcal{K}$ and $s = u+(s-u)$, so $u \in \mathcal{F}$ by (2.2). For any $\beta \in [0,1]$, $\beta u, (1-\beta)u \in \mathcal{K}$ and $\beta u + (1-\beta)u = u \in \mathcal{F}$, so by (2.2) we get $\beta u \in \mathcal{F}$. Similarly, for any $\beta > 1$, $\frac{1}{\beta}\beta u + \left(1-\frac{1}{\beta}\right)0 = u \in \mathcal{F}$ so $\beta u \in \mathcal{F}$. Hence (2.1) holds, i.e., $\mathcal{F}$ is a face of $\mathcal{K}$.

4. Let $x, y \in \mathcal{S}$ be such that $\alpha x + (1-\alpha)y \in \mathcal{F}_1 \cap \mathcal{F}_2$ for some $\alpha \in (0,1)$. By definition, for $i = 1, 2$, $\alpha x + (1-\alpha)y \in \mathcal{F}_i$ implies $x, y \in \mathcal{F}_i$. Hence $x, y \in \mathcal{F}_1 \cap \mathcal{F}_2$. This shows that $\mathcal{F}_1 \cap \mathcal{F}_2$ is a face of $\mathcal{S}$.

5. Suppose $x, y \in \mathcal{G}_1 \cap \mathcal{G}_2$ satisfy $\alpha x + (1-\alpha)y \in \mathcal{F}_1 \cap \mathcal{F}_2$ for some $\alpha \in (0,1)$. Then for $i = 1, 2$, $\alpha x + (1-\alpha)y \in \mathcal{F}_i$ and $x, y \in \mathcal{G}_i$ imply that $x, y \in \mathcal{F}_i$. Therefore $x, y \in \mathcal{F}_1 \cap \mathcal{F}_2$. This shows that $\mathcal{F}_1 \cap \mathcal{F}_2 \trianglelefteq \mathcal{G}_1 \cap \mathcal{G}_2$.

6. Let $x, y \in \mathcal{S}$ satisfy $z = \alpha x + (1-\alpha)y \in \mathcal{F}_1$ for some $\alpha \in (0,1)$. Then $z \in \mathcal{F}_2 \trianglelefteq \mathcal{S}$ so $x, y \in \mathcal{F}_2$. But $z \in \mathcal{F}_1 \trianglelefteq \mathcal{F}_1$ so $x, y \in \mathcal{F}_1$. This shows that $\mathcal{F}_1 \trianglelefteq \mathcal{S}$.

7. Let $x, y \in \mathcal{F}_2$ satisfy $\alpha x + (1-\alpha)y \in \mathcal{F}_1$ for some $\alpha \in (0,1)$. Since $\mathcal{F}_2 \subseteq \mathcal{S}$ and $\mathcal{F}_1 \trianglelefteq \mathcal{S}$, we get $x, y \in \mathcal{F}_1$.

8. If $\mathcal{F} = \mathcal{K} \cap \{a\}^{\perp}$ where $a \in \mathcal{K}^*$, then $\langle a, x\rangle_{\mathbb{V}} \geq 0$ for all $x \in \mathcal{K}$. Hence $\mathcal{F}$ is an exposed face of $\mathcal{K}$. Conversely, suppose that $\mathcal{F}$ is an exposed face of $\mathcal{K}$, and is exposed by the hyperplane $\{x : \langle a, x\rangle_{\mathbb{V}} = \alpha\}$. Then $0 \in \mathcal{F}$ implies that $\alpha = 0$ and $\mathcal{F} = \mathcal{K} \cap \{a\}^{\perp}$. Also, $\mathcal{K} \subseteq \{x : \langle a, x\rangle_{\mathbb{V}} \geq 0\}$ implies that $a \in \mathcal{K}^*$.

□

More results on faces of convex sets can be found in [6, 7, 8, 9, 50, 54] and [79, Chapter 18]. Here we recall the fact that, given a nonempty convex set S, the set of all relative interiors of the faces of S forms a partition of S itself.

Theorem 2.2.3 ([79], Theorem 18.2)**.** *Let C be a nonempty convex set and let U be the collection of all relative interiors of nonempty faces of C. Then U is a partition of C, i.e., the sets in U are disjoint and their union is C. Every relatively open convex subset of C is contained in one of the sets in U, and these are the maximal relatively open convex subsets of C.*

2.2.1 Minimal faces

Now we introduce the notion of minimal face.

Definition 2.2.4. *Let $\mathcal{K} \in \mathbb{V}$ be a nonempty convex cone and $\varnothing \neq \mathcal{S} \subseteq \mathcal{K}$. The minimal face of $\mathcal{K}$ containing $\mathcal{S}$ is defined as the set*

$$\operatorname{face}(\mathcal{S}, \mathcal{K}) := \bigcap \{\mathcal{F} : \mathcal{F} \trianglelefteq \mathcal{K},\ \mathcal{S} \subseteq \mathcal{F}\}.$$

Remark. In e.g. [13] and [81] for polyhedra, a minimal face of a convex set is defined as a face that does not contain any other face of that convex set.

We also caution that the notation we use for minimal faces may be used to mean different objects in the literature. In e.g. [62, 86], for any $Z \in \mathbb{S}^n_+$, $\operatorname{face}(\mathbb{S}^n_+, Z) := \left\{X \in \mathbb{S}^n_+ : \langle X, Z\rangle = 0\right\}$ is indeed the conjugate of the minimal face of $\mathbb{S}^n_+$ containing Z. (See Definition 2.2.7 for the definition of the conjugate face.)

By definition, $\operatorname{face}(\mathcal{S}, \mathcal{K})$ contains $\mathcal{S}$. By Item 4 of Proposition 2.2.2, $\operatorname{face}(\mathcal{S}, \mathcal{K})$ is indeed a face of $\mathcal{K}$.

There are a few equivalent conditions for a face of $\mathcal{K}$ to minimally contain a set $\mathcal{S} \subseteq \mathcal{K}$.

Proposition 2.2.5. *Let $\mathcal{K}$ be a nonempty convex cone, $\varnothing \neq \mathcal{S} \subseteq \mathcal{K}$, and $\mathcal{F} \trianglelefteq \mathcal{K}$. Then the following are equivalent:*

(1) $\mathcal{F} = \operatorname{face}(\mathcal{S}, \mathcal{K})$.

(2) $\mathcal{S} \subseteq \mathcal{F}$ *and* $\mathcal{S} \cap \operatorname{ri}(\mathcal{F}) \neq \varnothing$.

(3) $\operatorname{ri}(\operatorname{conv}(\mathcal{S})) \subseteq \operatorname{ri}(\mathcal{F})$.

In particular, $\operatorname{face}(\mathcal{S}, \mathcal{K}) = \operatorname{face}(s, \mathcal{K})$ *for any* $s \in \operatorname{ri}(\operatorname{conv}(\mathcal{S}))$.

Proof. We first show that (3) implies that $\mathcal{S} \subseteq \mathcal{F}$. Fix any $y \in \operatorname{ri}(\operatorname{conv}(\mathcal{S}))$. Let $x \in \mathcal{S}$. Then there exists $z \in \operatorname{conv}(\mathcal{S})$ such that $y = \alpha x + (1-\alpha)z \in \operatorname{ri}(\operatorname{conv}(\mathcal{S})) \subseteq \mathcal{F} \trianglelefteq \mathcal{K}$ for some $\alpha \in (0,1)$. Then $x \in \mathcal{F}$ by definition of faces. Hence $\mathcal{S} \subseteq \mathcal{F}$.

(3) $\implies$ (2): We already showed that $\mathcal{S} \subseteq \mathcal{F}$. Since $\mathcal{S} \neq \varnothing$, there exists $x \in \operatorname{ri}(\operatorname{conv}(\mathcal{S})) \subseteq \operatorname{ri}(\mathcal{F})$. Hence $\operatorname{conv}(\mathcal{S}) \cap \operatorname{ri}(\mathcal{F}) \neq \varnothing$.

(2) $\implies$ (1): Fix any $\hat{\mathcal{F}} \trianglelefteq \mathcal{K}$ with $\mathcal{S} \subseteq \hat{\mathcal{F}}$. We show that $\mathcal{F} \subseteq \hat{\mathcal{F}}$. Let $x \in \operatorname{conv}(\mathcal{S}) \cap \operatorname{ri}(\mathcal{F})$ and fix any $y \in \mathcal{F}$. Since $x \in \operatorname{ri}(\mathcal{F})$, there exists some $z \in \mathcal{F}$ such that $x = \alpha y + (1-\alpha)z$ for some $\alpha \in (0,1)$. But $y, z \in \mathcal{K}$, so $x \in \mathcal{S} \subseteq \hat{\mathcal{F}} \trianglelefteq \mathcal{K}$ implies $y, z \in \hat{\mathcal{F}}$. Hence $\mathcal{F} \subseteq \hat{\mathcal{F}}$, and $\operatorname{face}(\mathcal{S}, \mathcal{K}) = \mathcal{F}$.

(1) $\implies$ (3): We show that $\operatorname{ri}(\operatorname{conv}(\mathcal{S})) \subseteq \operatorname{ri}(\operatorname{face}(\mathcal{S}, \mathcal{K}))$. By Theorem 2.2.3, there exists a unique $\hat{\mathcal{F}} \trianglelefteq \mathcal{K}$ such that $\operatorname{ri}(\operatorname{conv}(\mathcal{S})) \subseteq \operatorname{ri}(\hat{\mathcal{F}})$. We claim that $\hat{\mathcal{F}} = \operatorname{face}(\mathcal{S}, \mathcal{K})$. First, $\operatorname{ri}(\operatorname{conv}(\mathcal{S})) \subseteq \operatorname{ri}(\hat{\mathcal{F}})$ implies $\mathcal{S} \subseteq \hat{\mathcal{F}}$, so $\operatorname{face}(\mathcal{S}, \mathcal{K}) \subseteq \hat{\mathcal{F}}$. Now suppose that $\tilde{\mathcal{F}} \trianglelefteq \mathcal{K}$ contains $\mathcal{S}$. We show that any $x \in \hat{\mathcal{F}}$ lies in $\tilde{\mathcal{F}}$. Indeed, let $y \in \operatorname{ri}(\operatorname{conv}(\mathcal{S})) \subseteq \operatorname{ri}(\hat{\mathcal{F}})$. Then there exist $z \in \hat{\mathcal{F}}$ and $\alpha \in (0,1)$ such that $y = \alpha x + (1-\alpha)z$. But $y \in \tilde{\mathcal{F}}$, so by definition of faces we get $x \in \tilde{\mathcal{F}}$ too. Consequently, we get $\operatorname{ri}(\operatorname{conv}(\mathcal{S})) \subseteq \operatorname{ri}(\operatorname{face}(\mathcal{S}, \mathcal{K}))$.

Finally, for any $s \in \operatorname{ri}(\operatorname{conv}(\mathcal{S}))$, Item (3) indicates that $s \in \operatorname{ri}(\operatorname{face}(\mathcal{S}, \mathcal{K}))$, so by Item (2) we have $\operatorname{face}(s, \mathcal{K}) = \operatorname{face}(\mathcal{S}, \mathcal{K})$. □

An immediate consequence of Proposition 2.2.5 is that, if a set $\mathcal{S} \subseteq \mathcal{K}$ contains a relative interior point of $\mathcal{K}$, then the minimal face containing $\mathcal{S}$ has to be the entire cone $\mathcal{K}$.

Corollary 2.2.6. *Let $\mathcal{K}$ be a convex cone and $\mathcal{F} \trianglelefteq \mathcal{K}$. Then $\mathcal{F} \cap \operatorname{ri}(\mathcal{K}) \neq \varnothing$ if and only if $\mathcal{F} = \mathcal{K}$.*

Proof. By Proposition 2.2.5, $\mathcal{F} \cap \operatorname{ri}(\mathcal{K}) \neq \varnothing$ implies $\mathcal{K} = \operatorname{face}(\mathcal{F}, \mathcal{K}) = \mathcal{F}$. Conversely, if $\mathcal{F} = \mathcal{K}$, then $\mathcal{F} \cap \operatorname{ri}(\mathcal{K}) = \operatorname{ri}(\mathcal{K}) \neq \varnothing$. □

2.2.2 Conjugate faces

In linear algebra, given a linear subspace $\mathcal{L}$ lying in an inner product space, we can consider the orthogonal subspace $\mathcal{L}^\perp$ of $\mathcal{L}$. Similarly, given a face $\mathcal{F}$ of a closed convex cone, we can consider a particular face of that cone orthogonal to $\mathcal{F}$, which we call the conjugate face.

Definition 2.2.7. *Let $\mathcal{K}$ be a closed convex cone and let $\mathcal{F}$ be a face of $\mathcal{K}$. The conjugate face of $\mathcal{K}$ with respect to $\mathcal{F}$ is defined as $\mathcal{F}^c := \mathcal{F}^\perp \cap \mathcal{K}^*$.*

It is easy to check that $\mathcal{F}^\perp \cap \mathcal{K}^*$ is indeed a face of $\mathcal{K}^*$ if $\mathcal{F} \trianglelefteq \mathcal{K}$. In fact, for any nonempty set $\mathcal{S} \subseteq \mathcal{K}$, the set $\mathcal{S}^\perp \cap \mathcal{K}^*$ is a face of $\mathcal{K}^*$.

Proposition 2.2.8. *Let $\varnothing \neq \mathcal{S} \subseteq \mathcal{K}$. Then for any $z \in \mathrm{ri}(\mathrm{conv}(\mathcal{S}))$, $\mathcal{S}^\perp \cap \mathcal{K}^* = \{z\}^\perp \cap \mathcal{K}^*$ is a face of $\mathcal{K}^*$.*

Proof. Since $\mathcal{S}^\perp$ is a linear subspace of $\mathbb{V}$, $\mathcal{S}^\perp \cap \mathcal{K}^*$ is a closed convex cone. Moreover, $z \in \mathcal{S}$ implies $\mathcal{S}^\perp \cap \mathcal{K}^* \subseteq \{z\}^\perp \cap \mathcal{K}^*$.

Suppose that $u \in \{z\}^\perp \cap \mathcal{K}^*$. We show that $u \in \mathcal{S}^\perp \cap \mathcal{K}^*$. Fix any $\tilde{z} \in \mathcal{S}$. Then there exists $\alpha > 1$ such that $(1-\alpha)\tilde{z} + \alpha z \in \mathrm{conv}(\mathcal{S}) \subseteq \mathcal{K}$. Then $0 \leq \langle (1-\alpha)\tilde{z} + \alpha z, u\rangle_{\mathbb{V}} = (1-\alpha)\langle \tilde{z}, u\rangle_{\mathbb{V}} \leq 0$ since $\alpha > 1$. This means $u \in \mathcal{S}^\perp$. Hence $\mathcal{S}^\perp \cap \mathcal{K}^* = \{z\}^\perp \cap \mathcal{K}^*$.

To see that $\mathcal{S}^\perp \cap \mathcal{K}^*$ is a face of $\mathcal{K}^*$, suppose that $x, y \in \mathcal{K}^*$ satisfy $u := x+y \in \{z\}^\perp \cap \mathcal{K}^*$. Since $z \in \mathcal{K}$, we have $\langle z, x\rangle_{\mathbb{V}} \geq 0$ and $\langle z, y\rangle_{\mathbb{V}} \geq 0$. Then $\langle z, x+y\rangle_{\mathbb{V}} = 0$ implies $\langle z, x\rangle_{\mathbb{V}} = 0 = \langle z, y\rangle_{\mathbb{V}}$. Hence $x, y \in \{z\}^\perp \cap \mathcal{K}^*$. This shows that $\mathcal{S}^\perp \cap \mathcal{K}^* = \{z\}^\perp \cap \mathcal{K}^*$ is a face of $\mathcal{K}^*$. □

An immediate consequence of Proposition 2.2.8 is that the conjugate face of any face of a convex cone is exposed. In fact:

Corollary 2.2.9. *[91, Proposition 3.1, Item 2] Let $\mathcal{K}$ be a closed convex cone and let $\mathcal{F} \trianglelefteq \mathcal{K}$. Then $\mathcal{F}^c$ is an exposed face of $\mathcal{K}^*$. Moreover, $\mathcal{F}^{cc} := (\mathcal{F}^c)^c = \mathcal{F}$ if and only if $\mathcal{F}$ is an exposed face of $\mathcal{K}$.*

Proof. Let $x \in \mathrm{ri}(\mathcal{F})$. Then $\mathcal{F}^c = \{x\}^\perp \cap \mathcal{K}^*$ is an exposed face of $\mathcal{K}^*$ by Proposition 2.2.8.

If $\mathcal{F} = \mathcal{F}^{cc}$, then $\mathcal{F}$ as a conjugate face is exposed. Conversely, suppose that $\mathcal{F} = \{x\}^\perp \cap \mathcal{K}$ for some $0 \neq x \in \mathcal{K}^*$. Then $\mathcal{F} \subseteq \{x\}^\perp$ implies $x \in \mathcal{F}^\perp$, so $0 \neq x \in \mathcal{F}^c$. Fix any $0 \neq y \in \mathrm{ri}(\mathcal{F}^c)$. Then $y - \epsilon x \in \mathcal{F}^c$ for some small $\epsilon > 0$. If $z \in \mathcal{K} \cap \{y\}^\perp$, then $y - \epsilon x \in \mathcal{K}^*$ implies that $0 \leq \langle z, y - \epsilon x\rangle_{\mathbb{V}} = -\epsilon\langle z, x\rangle_{\mathbb{V}} \leq 0$, i.e., $\langle z, x\rangle_{\mathbb{V}} = 0$. Hence $\mathcal{K} \cap \{y\}^\perp \subseteq \mathcal{F}$. If $z \in \mathcal{F}$, then $y \in \mathcal{F}^\perp \cap \mathcal{K}^*$ implies that $\langle z, y\rangle_{\mathbb{V}} = 0$. Hence $\mathcal{F} \subseteq \mathcal{K} \cap \{y\}^\perp$. Therefore $\mathcal{F} = \mathcal{K} \cap \{y\}^\perp$ and $0 \neq y \in \mathcal{F}^c \subseteq \mathcal{K}^*$, i.e., $\mathcal{F}$ is an exposed face. □

Another consequence of Proposition 2.2.8 is that for any nonempty convex set $\mathcal{S} \subseteq \mathcal{K}$, $\mathcal{S}^\perp \cap \mathcal{K}^*$ is the conjugate face of the minimal face of $\mathcal{K}$ containing $\mathcal{S}$.

Corollary 2.2.10. *Let $\varnothing \neq \mathcal{S} \subseteq \mathcal{K}$. Then* $(\mathrm{face}(\mathcal{S}, \mathcal{K}))^c = \mathcal{S}^\perp \cap \mathcal{K}^*$. *Moreover, if* $\mathrm{face}(\mathcal{S}, \mathcal{K})$ *is an exposed face, then* $\mathrm{face}(\mathcal{S}, \mathcal{K}) = (\mathcal{S}^\perp \cap \mathcal{K}^*)^c$.

Proof. By Proposition 2.2.5, $\mathrm{ri}(\mathrm{conv}(\mathcal{S})) \subseteq \mathrm{ri}\,(\mathrm{face}(\mathcal{S}, \mathcal{K}))$, so for any $z \in \mathrm{ri}(\mathrm{conv}(\mathcal{S}))$,

$$\mathcal{S}^\perp \cap \mathcal{K}^* = \{z\}^\perp \cap \mathcal{K}^* = (\mathrm{face}(\mathcal{S}, \mathcal{K}))^\perp \cap \mathcal{K}^* = (\mathrm{face}(\mathcal{S}, \mathcal{K}))^c\,,$$

by Proposition 2.2.8. The second claim follows from Corollary 2.2.9. □

One way of identifying a smaller face of $\mathcal{K}$ containing the feasible region $\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}$ of the conic program ($\mathrm{P}_{\mathrm{conic}}$) is to find a *direction of constancy* $d \in \mathcal{K}^*$ for the dual ($\mathrm{D}_{\mathrm{conic}}$) and take the conjugate face $\{d\}^\perp \cap \mathcal{K}$, which Proposition 2.2.8 has shown to be a face of $\mathcal{K}$. (See Section 4.1.) More details on conjugate faces can be found in [91].

2.2.3 Examples

In this section we characterize the faces of two well-studied cones, the second-order cone $\mathcal{Q}^n$ and the positive semidefinite cone $\mathbb{S}^n_+$.

Faces of $\mathcal{Q}^n$

In this section, we will show that all the proper faces of $\mathcal{Q}^n$ are of the form $\left\{ \alpha \begin{pmatrix} \|x\| \\ x \end{pmatrix} : \alpha \geq 0 \right\}$, where $x \in \mathbb{R}^{n-1}$. In other words, all faces of $\mathcal{Q}^n$ are extreme rays of dimension 1.

Lemma 2.2.11. *Let $x \in \mathbb{R}^{n-1}$. Then the set $\mathcal{F} := \left\{ \alpha \begin{pmatrix} \|x\| \\ x \end{pmatrix} : \alpha \geq 0 \right\}$ is a face of $\mathcal{Q}^n$.*

Proof. First note that $\mathcal{F}$ is a closed convex cone in $\mathbb{R}^n$ and a subset of $\mathcal{Q}^n$. Let $y, z \in \mathcal{Q}^n$ satisfy $y + z \in \mathcal{F}$, i.e.,

$$y_1 \geq \|y_{2:n}\|, \quad z_1 \geq \|z_{2:n}\|, \quad y_1 + z_1 = \alpha\|x\|, \quad y_{2:n} + z_{2:n} = \alpha x \tag{2.3}$$

for some $\alpha \geq 0$. We show that $y, z \in \mathcal{F}$.

If $\alpha\|x\| = 0$, then $y_1 + z_1 = 0$, implying $y_1 = 0 = z_1$. Hence $y = z = 0 \in \mathcal{F}$.

Suppose that $\alpha\|x\| > 0$. Since $\alpha \geq 0$,

$$\alpha\|x\| = \|y_{2:n} + z_{2:n}\| \leq \|y_{2:n}\| + \|z_{2:n}\| \leq y_1 + z_1 = \alpha\|x\|,$$

implying

$$0 < \|y_{2:n} + z_{2:n}\| = \|y_{2:n}\| + \|z_{2:n}\|, \quad \|y_{2:n}\| = y_1, \quad \text{and} \quad \|z_{2:n}\| = z_1. \tag{2.4}$$

Without loss of generality, assume $y_{2:n} \neq 0$. Then (2.4) implies that $z_{2:n} = \xi y_{2:n}$ for some $\xi \neq -1$ (otherwise $y_{2:n} + z_{2:n} = 0$, contradicting $\|y_{2:n} + z_{2:n}\| > 0$). It follows from (2.3) and (2.4) that

$$\alpha\|x\| = y_1 + z_1 = \|y_{2:n}\| + \|z_{2:n}\| = (1 + |\xi|)\|y_{2:n}\|, \quad \text{and} \quad \alpha x = y_{2:n} + z_{2:n} = (1 + \xi)y_{2:n}.$$

Therefore $(1+|\xi|)\|y_{2:n}\| = |1+\xi|\|y_{2:n}\|$, implying that $\xi \geq 0$.[2] Hence (2.4) implies that $z_1 = |\xi y_1| = \xi y_1$, and

$$y = \frac{\alpha}{1+\xi}x,\ z = \xi y = \frac{\alpha\xi}{1+\xi}x \in \mathcal{F}.$$

This shows that $\mathcal{F} \lhd \mathcal{Q}^n$. □

So we see that sets like $\mathcal{F}$ are extreme rays of $\mathcal{Q}^n$. Before we show that they form all the proper faces of $\mathcal{Q}^n$, we prove the following simple lemma.

Lemma 2.2.12. *Let $\mathcal{F} \trianglelefteq \mathcal{Q}^n$ contain $\begin{pmatrix}\alpha\\ x\end{pmatrix}, \begin{pmatrix}\beta\\ y\end{pmatrix}$ with $x, y \in \mathbb{R}^{n-1}$ being linearly independent. Then $\mathcal{F} = \mathcal{Q}^n$.*

Proof. Note that since x and y are linearly independent,

$$\|x+y\| < \|x\| + \|y\| \leq \alpha + \beta.$$

Then $\begin{pmatrix}\alpha+\beta\\ x+y\end{pmatrix} \in \mathcal{F} \cap \operatorname{int}(\mathcal{Q}^n)$. By Corollary 2.2.6, we get $\mathcal{F} = \mathcal{Q}^n$. □

We summarize the results of Lemmas 2.2.11 and 2.2.12 for the characterization of faces of $\mathcal{Q}^n$.

Theorem 2.2.13. *Let $\mathcal{F} \subseteq \mathcal{Q}^n$. Then $\mathcal{F} \trianglelefteq \mathcal{Q}^n$ if and only if $\mathcal{F} = \left\{\alpha\begin{pmatrix}\|x\|\\ x\end{pmatrix} : \alpha \geq 0\right\}$ for some $x \in \mathbb{R}^{n-1}$ or $\mathcal{F} = \mathcal{Q}^n$.*

Proof. We already saw from Lemma 2.2.11 that if $\mathcal{F} = \left\{\alpha\begin{pmatrix}\|x\|\\ x\end{pmatrix} : \alpha \geq 0\right\}$ for some $x \in \mathbb{R}^{n-1}$, then $\mathcal{F} \trianglelefteq \mathcal{Q}^n$. For the converse, suppose that $\mathcal{F} \trianglelefteq \mathcal{Q}^n$. By Lemma 2.2.12, either

(1) $\mathcal{F} = \mathcal{Q}^n$, or

(2) $\mathcal{F} = \{0\}$, or

(3) $\{0\} \neq \mathcal{F} \subsetneq \mathcal{Q}^n$ and for any $y, z \in \mathcal{F}$, $y_{2:n}$ and $z_{2:n}$ are linearly dependent.

We first show that if $\mathcal{F} \lhd \mathcal{Q}^n$, then for any $z \in \mathcal{F}$, $z_1 = \|z_{2:n}\|$. If not, then $z_1 > \|z_{2:n}\|$ and for any $0 \neq w \in \mathcal{Q}^n$, $z \pm \alpha w \in \mathcal{Q}^n$ for any $\alpha \in \left(0, \frac{z_1 - \|z_{2:n}\|}{w_1 + \|w_{2:n}\|}\right)$. Then $\frac{1}{2}(y+\alpha w) + \frac{1}{2}(y-\alpha w) = y \in \mathcal{F} \lhd \mathcal{Q}^n$

[2] If $1+\xi < 0$, then $1+|\xi| = -1-\xi$. If $\xi < 0$, then we get $1 = -1$ which is impossible; if $\xi \geq 0$, we get $2\xi = -2$, which contradicts the fact that $\xi \neq -1$. Hence we must have $1+\xi \geq 0$. This implies that $|\xi| = \xi$, i.e, $\xi \geq 0$.

implies that $y \in \alpha w \in \mathcal{F}$. But $y, \alpha w \in \mathcal{Q}^n$ and $\mathcal{Q}^n$ is a cone, so $w \in \mathcal{F}$. Therefore $\mathcal{Q}^n \subseteq \mathcal{F}$, which is contradictory. This shows that for any $z \in \mathcal{F}$, $z_1 = \|z_{2:n}\|$.

Now consider Case (3); pick any $0 \neq z \in \mathcal{F}$. Since $z \neq 0$ and $z_1 = \|z_{2:n}\|$, we have that $z_{2:n} \neq 0$. We prove that $\mathcal{F} = \{\alpha z : \alpha \geq 0\}$. Since $z \in \mathcal{F}$, $\{\alpha z : \alpha \geq 0\} \subseteq \mathcal{F}$. Conversely, for any $w \in \mathcal{F}$, we have $w_{2:n} = \alpha z_{2:n}$ for some $\alpha \in \mathbb{R}$. By Lemma 2.2.12, $w_1 = \|w_{2:n}\| = |\alpha|\|z_{2:n}\| = |\alpha| z_1$. If $\alpha < 0$, then $z \in \mathcal{F}$ implies $-\alpha z \in \mathcal{F}$. Hence

$$\mathcal{F} \ni w - \alpha z = \begin{pmatrix} -\alpha z_1 \\ \alpha z_{2:n} \end{pmatrix} - \alpha z = \begin{pmatrix} -2\alpha z_1 \\ 0 \end{pmatrix},$$

but $-2\alpha z_1 > 0$, which is absurd. Hence we must have $\alpha \geq 0$. This implies that $w = \alpha z$. Therefore we get $\mathcal{F} = \{\alpha z : \alpha \geq 0\} = \left\{ \alpha \begin{pmatrix} \|z_{2:n}\| \\ z_{2:n} \end{pmatrix} : \alpha \geq 0 \right\}$. □

Faces of $\mathbb{S}^n_+$

In this section we

- characterize all the proper faces of $\mathbb{S}^n_+$,
- show that all faces of $\mathbb{S}^n_+$ are exposed faces, and
- compute the conjugate faces (i.e., the sets $\mathbb{S}^n_+ \cap \{D\}^\perp$, where $D \in \mathbb{S}^n_+$) of $\mathbb{S}^n_+$.

Proposition 2.2.14. *A nonempty set $\mathcal{F} \subset \mathbb{S}^n_+$ is a face of $\mathbb{S}^n_+$ if and only if $\mathcal{F} = \{0\}$ or $\mathcal{F} = Q\mathbb{S}^r_+Q^\top$ for some $Q \in \mathbb{R}^{n\times r}$ and $0 < r \leq n$.*

Proof. Recalling that $\mathbb{S}^n_+$ is a pointed cone,[3] the set $\{0\}$ is a proper face of $\mathbb{S}^n_+$ by Item 1 of Proposition 2.2.2.

We show that $Q\mathbb{S}^r_+Q^\top$ is a face of $\mathbb{S}^n_+$ for any $Q \in \mathbb{R}^{n\times r}$. For any $X, Y \in \mathbb{S}^r_+$ and $\alpha \geq 0$, $\alpha QXQ^\top + QYQ^\top = Q(\alpha X + Y)Q^\top$; hence $Q\mathbb{S}^r_+Q^\top$ is a convex cone. To see that $Q\mathbb{S}^r_+Q^\top$ is indeed a face of $\mathbb{S}^n_+$, let $X, Y \in \mathbb{S}^n_+$ with $X + Y = QWQ^\top$ for some $W \in \mathbb{S}^r_+$. Let P be a full column rank matrix such that $\ker(Q^\top) = \mathrm{range}(P)$. Then $Q^\top P = 0$, and $\langle X + Y, PP^\top \rangle = \langle QWQ^\top, PP^\top \rangle = 0$ implies that $\langle X, PP^\top \rangle = 0 = \langle Y, PP^\top \rangle$ as $X, Y, PP^\top \succeq 0$. Then $\mathrm{range}(X) \subseteq \ker(P^\top) = \mathrm{range}(Q)$. This together with $X \succeq 0$ implies that $X \in Q\mathbb{S}^r_+Q^\top$. Similarly, $Y \in Q\mathbb{S}^r_+Q^\top$.

Now we show that any nonzero face $\mathcal{F}$ of $\mathbb{S}^n_+$ equals to $Q\mathbb{S}^r_+Q^\top$ for some $Q \in \mathbb{R}^{n\times r}$. Let $X \in \mathrm{ri}(\mathcal{F})$, and $X = QD_+Q^\top$ be the compact spectral decomposition of X (with $Q \in \mathbb{R}^{n\times r}$ and

[3] If $X \in \mathbb{S}^n$ satisfies $X \succeq 0$ and $-X \succeq 0$, then we must have $\lambda_{\max}(X) = \lambda_{\min}(X) = 0$. This implies that $X = 0$.

$D_+ \in \mathbb{S}^r_{++}$). We claim that $\mathcal{F} = Q\mathbb{S}^r_+Q^\top$. To see that any $Q\bar{Y}Q^\top \in \mathcal{F}$ for any $\bar{Y} \in \mathbb{S}^r_+$, note that $D_+ \succ 0$ so we have $D_+ - \alpha\bar{Y} \succeq 0$ for sufficiently small $\alpha \in (0,1)$. Since $D_+ = (D_+ - \alpha\bar{Y}) + \alpha\bar{Y}$ and $QD_+Q^\top \in \mathcal{F}$, we have $Q(\alpha\bar{Y})Q^\top \in \mathcal{F}$. Hence $Q\bar{Y}Q^\top \in \mathcal{F}$. Conversely, fix any $Y \in \mathcal{F}$. Since $X \in \mathrm{ri}(\mathcal{F})$, $(1+\epsilon)X - \epsilon Y \in \mathcal{F}$ for some small $\epsilon > 0$. Let $P \in \mathbb{R}^{n\times(n-r)}$ satisfy $\ker(Q^\top) = \mathrm{range}(P)$. Then $0 \leq \langle PP^\top, (1+\epsilon)X - \epsilon Y\rangle = -\epsilon\langle PP^\top, Y\rangle \leq 0$, implying that $PP^\top Y = 0$. Therefore $\mathrm{range}(Y) \subseteq \ker(P^\top) = \mathrm{range}(Q)$, implying that $Y \in Q\mathbb{S}^r_+Q^\top$. □

Remark. In Proposition 2.2.14, the matrix Q does not have to be of full column rank; nonetheless, in practice we often pick Q that is of full column rank, or even Q with orthonormal columns. In fact, let $Q_{\mathrm{QR}} \in \mathbb{R}^{n\times\mathrm{rank}(Q)}$ have orthonormal columns and satisfy $\mathrm{range}(Q_{\mathrm{QR}}) = \mathrm{range}(Q)$; then $Q\mathbb{S}^r_+Q^\top = Q_{\mathrm{QR}}\mathbb{S}^{\mathrm{rank}(Q)}_+Q_{\mathrm{QR}}^\top$.

A direct result of Proposition 2.2.14 is that all the faces of $\mathbb{S}^n_+$ are exposed, i.e., they are all of form $\mathbb{S}^n_+ \cap \{X\}^\perp$ for some $X \in \mathbb{S}^n_+$.

Corollary 2.2.15. *Let $\mathcal{F} \subseteq \mathbb{S}^n_+$. Then $\mathcal{F}$ is a face of $\mathbb{S}^n_+$ if and only if $\mathcal{F} = \mathbb{S}^n_+ \cap \{X\}^\perp$ for some $X \in \mathbb{S}^n_+$, and $\mathcal{F}$ is a proper face of $\mathbb{S}^n_+$ if and only if $\mathcal{F} = \mathbb{S}^n_+ \cap \{X\}^\perp$ for some nonzero $X \in \mathbb{S}^n_+$.*

Proof. We first show that for any $X \in \mathbb{S}^n_+$, $\mathbb{S}^n_+ \cap \{X\}^\perp$ is a face of $\mathbb{S}^n_+$. If $X = 0$, then $\mathbb{S}^n_+ \cap \{X\}^\perp = \mathbb{S}^n_+$. If $X \in \mathbb{S}^n_{++}$, then $\mathbb{S}^n_+ \cap \{X\}^\perp = \{0\}$ is a proper face of $\mathbb{S}^n_+$. Now suppose that $X \neq 0$ is not positive definite. Let $X = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P^\top \\ Q^\top \end{bmatrix}$, where $\begin{bmatrix} P & Q \end{bmatrix} \in \mathbb{R}^{n\times n}$ is an orthogonal matrix and $D_+ \in \mathbb{S}^{n-r}_+$. Then $\mathbb{S}^n_+ \cap \{X\}^\perp = Q\mathbb{S}^r_+Q^\top$, which is a proper face of $\mathbb{S}^n_+$ by Proposition 2.2.14.

Conversely, observe that $\mathbb{S}^n_+ = \mathbb{S}^n_+ \cap \{0\}^\perp$ and $\{0\} = \mathbb{S}^n_+ \cap \{I\}^\perp$. Let $\mathcal{F} \trianglelefteq \mathbb{S}^n_+$ be a proper face. By Proposition 2.2.14, $\mathcal{F} = Q\mathbb{S}^r_+Q^\top$ for some $Q \in \mathbb{S}^r_+$ with orthonormal columns. Let $\begin{bmatrix} P & Q \end{bmatrix} \in \mathbb{R}^{n\times n}$ is an orthogonal matrix. Then $PP^\top \in \mathbb{S}^n_+$ and $\mathcal{F} = \mathbb{S}^n_+ \cap \left\{PP^\top\right\}^\perp$. □

Another consequence of Proposition 2.2.14 is that for any nonempty $\mathcal{S} \subseteq \mathbb{S}^n_+$ and any invertible matrix $Q \in \mathbb{R}^{n\times n}$, the minimal face of $\mathcal{S}$ is "unchanged" under the conjugation by $Q \cdot Q^\top$ in the following sense:

Corollary 2.2.16. *Let $\varnothing \neq \mathcal{S} \subseteq \mathbb{S}^n_+$ and $Q \in \mathbb{R}^{n\times n}$ be invertible. Then $Q(\mathrm{face}(\mathcal{S}, \mathbb{S}^n_+))Q^\top = \mathrm{face}(Q\mathcal{S}Q^\top, \mathbb{S}^n_+)$.*

Proof. First observe that for any nonempty convex set $\mathcal{F} \subseteq \mathbb{S}^n_+$, $\mathcal{F} \trianglelefteq \mathbb{S}^n_+$ if and only if $Q\mathcal{F}Q^\top \trianglelefteq \mathbb{S}^n_+$. In fact, if $\mathcal{F} \trianglelefteq \mathbb{S}^n_+$, then $\mathcal{F} = \mathbb{S}^n_+$ or $\{0\}$, or $\mathcal{F} = P\mathbb{S}^r_+P^\top$ for some $P \in \mathbb{R}^{n\times r}$ and $0 < r < n$.

Then $Q\mathcal{F}Q^\top = \mathbb{S}^n_+$ or $\{0\}$ or $(QP)\mathbb{S}^r_+(QP)^\top$, so by Proposition 2.2.14 $Q\mathcal{F}Q^\top \trianglelefteq \mathbb{S}^n_+$. Conversely, if $Q\mathcal{F}Q^\top \trianglelefteq \mathbb{S}^n_+$, then $\mathcal{F} = Q^{-1}(Q\mathcal{F}Q^\top)Q^{-\top} \trianglelefteq \mathbb{S}^n_+$.

Now observe that

$$\begin{aligned}
Q(\text{face}(\mathcal{S}, \mathbb{S}^n_+))Q^\top &= \left\{QXQ^\top \in \mathbb{S}^n : X \in \mathcal{F},\ \forall\, \mathcal{F} \trianglelefteq \mathbb{S}^n_+ \text{ s.t. } \mathcal{S} \subseteq \mathcal{F}\right\} \\
&= \left\{Y \in \mathbb{S}^n : Y \in Q\mathcal{F}Q^\top,\ \forall\, \mathcal{F} \trianglelefteq \mathbb{S}^n_+ \text{ s.t. } \mathcal{S} \subseteq \mathcal{F}\right\} \\
&= \left\{Y \in \mathbb{S}^n : Y \in \mathcal{F},\ \forall\, \mathcal{F} \trianglelefteq \mathbb{S}^n_+ \text{ s.t. } Q\mathcal{S}Q^\top \subseteq \mathcal{F}\right\} \\
&= \text{face}(Q\mathcal{S}Q^\top, \mathbb{S}^n_+).
\end{aligned}$$

Therefore $Q(\text{face}(\mathcal{S}, \mathbb{S}^n_+))Q^\top = \text{face}(Q\mathcal{S}Q^\top, \mathbb{S}^n_+)$. □

In fact, the result of Corollary 2.2.16 also holds if Q has orthonormal columns.

Proposition 2.2.17. *Let $\varnothing \neq \mathcal{S} \subseteq \mathbb{S}^r_+$ and $Q \in \mathbb{R}^{n\times r}$ have orthonormal columns. Then $Q(\text{face}(\mathcal{S}, \mathbb{S}^r_+))Q^\top = \text{face}(Q\mathcal{S}Q^\top, \mathbb{S}^n_+)$.*

Proof. By Proposition 2.2.5, there exist $\epsilon > 0$ and $X \in \mathcal{S}$ such that

$$\{Y \in \mathbb{S}^r : \|Y - X\|_F \leq \epsilon\} \cap \text{aff}(\text{face}(\mathcal{S}, \mathbb{S}^r_+)) \ \subseteq\ \text{face}(\mathcal{S}, \mathbb{S}^r_+). \tag{2.5}$$

Let $\hat{X} := QXQ^\top \in Q\mathcal{S}Q^\top$. We show that $\hat{X} \in \text{ri}\left(Q(\text{face}(\mathcal{S}, \mathbb{S}^r_+))Q^\top\right)$. Indeed, pick any $\hat{Y} \in \mathbb{S}^n$ satisfying

$$\|\hat{Y} - \hat{X}\|_F \leq \epsilon, \quad \hat{Y} \in \text{aff}\left(Q(\text{face}(\mathcal{S}, \mathbb{S}^r_+))Q^\top\right) = Q\left(\text{aff}(\text{face}(\mathcal{S}, \mathbb{S}^r_+))\right)Q^\top. \tag{2.6}$$

Then $\hat{Y} = QYQ^\top$ for some $Y \in \text{aff}(\text{face}(\mathcal{S}, \mathbb{S}^r_+))$. Moreover, $\|Y - X\|_F = \|\hat{Y} - \hat{X}\|_F \leq \epsilon$. By (2.5), we get $Y \in \text{face}(\mathcal{S}, \mathbb{S}^r_+)$. Therefore $\hat{Y} = QYQ^\top \in Q(\text{face}(\mathcal{S}, \mathbb{S}^r_+))Q^\top$. Since $\hat{Y}$ satisfying (2.6) is arbitrary, we get that $\hat{X} \in \text{ri}\left(Q(\text{face}(\mathcal{S}, \mathbb{S}^r_+))Q^\top\right) \cap Q\mathcal{S}Q^\top$. Since $Q(\text{face}(\mathcal{S}, \mathbb{S}^r_+))Q^\top$ is a face of $\mathbb{S}^n_+$, by Proposition 2.2.5 we get that $Q(\text{face}(\mathcal{S}, \mathbb{S}^r_+))Q^\top = \text{face}(Q\mathcal{S}Q^\top, \mathbb{S}^n_+)$. □

Using the fact that all the nonzero faces of $\mathbb{S}^n_+$ are of the form $Q\mathbb{S}^r_+Q^\top$ for some full rank matrix $Q \in \mathbb{R}^{n\times r}$, we can prove a simple result that any two matrices in the relative interior of a set $\mathcal{S} \subseteq \mathbb{S}^n_+$ have the same rank.

Corollary 2.2.18. *Let $\mathcal{S} \subseteq \mathbb{S}^n_+$ be nonempty convex. Then all matrices in* $\text{ri}(\mathcal{S})$ *have the same rank and are the maximum-rank elements of $\mathcal{S}$.*

Proof. The claim immediately holds if $\mathcal{S} = \{0\}$. Suppose that $\mathcal{S} \neq \{0\}$. By Proposition 2.2.14, there exists a full column rank matrix $Q \in \mathbb{R}^{n\times r}$ (with $r \leq n$) such that $\text{face}(\mathcal{S}, \mathbb{S}^n_+) = Q\mathbb{S}^r_+Q^\top$. In particular, any matrix in $\mathcal{S}$ is of rank at most r. Now fix any $X, Y \in \text{ri}(\mathcal{S})$. By Proposition 2.2.5, $X, Y \in \text{ri}(Q\mathbb{S}^r_+Q^\top) = Q\mathbb{S}^r_{++}Q^\top$. Hence both X, Y are of rank r. □

Remark. The maximum rank of elements in a convex subset $\mathcal{S}$ of $\mathbb{S}^n_+$ can be found using the information on the minimal face of $\mathbb{S}^n_+$ containing $\mathcal{S}$. Also, all the matrices in the relative interior of the set of feasible/optimal solutions of an SDP (to be introduced in the next chapter) have the same rank.

Now we give the explicit expression of the conjugate face and the dual cone of a proper face $\mathcal{F} = Q\mathbb{S}^r_+Q^\top \trianglelefteq \mathbb{S}^n_+$, where $0 < r < n$ and $\begin{bmatrix} P & Q \end{bmatrix} \in \mathbb{R}^{n\times n}$ is an orthogonal matrix. Indeed,

$$(Q\mathbb{S}^r_+Q^\top)^c = P\mathbb{S}^{n-r}_+P^\top \quad \text{and} \quad (Q\mathbb{S}^r_+Q^\top)^* = \left\{X \in \mathbb{S}^n : Q^\top XQ \in \mathbb{S}^r_+\right\}.$$

Faces of Cartesian products of convex cones

This example is a brief discussion on the Cartesian products of convex cones, which are often encountered in practice. We take note of the fact that the faces of a Cartesian product of pointed convex cones are the same as the Cartesian products of the faces of those convex cones.

Proposition 2.2.19. *Let $\mathcal{K}_1$, $\mathcal{K}_2$ be convex cones containing 0. Then $\mathcal{F} \trianglelefteq \mathcal{K}_1 \times \mathcal{K}_2$ if and only if $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$ for some faces $\mathcal{F}_1 \trianglelefteq \mathcal{K}_1$ and $\mathcal{F}_2 \trianglelefteq \mathcal{K}_2$.*

Proof. Suppose $\mathcal{F} \trianglelefteq \mathcal{K}_1 \times \mathcal{K}_2$. Define

$$\mathcal{F}_1 := \left\{x^{(1)} \in \mathcal{K}_1 : (x^{(1)}, 0) \in \mathcal{F}\right\},$$
$$\mathcal{F}_2 := \left\{x^{(2)} \in \mathcal{K}_2 : (0, x^{(2)}) \in \mathcal{F}\right\}.$$

Then $\mathcal{F}_1$ and $\mathcal{F}_2$ are convex sets and contain 0 by Item 1 of Proposition 2.2.2. We show that $\mathcal{F}_1 \trianglelefteq \mathcal{K}_1$. Let $x^{(1)}, y^{(1)} \in \mathcal{K}_1$ satisfy $x^{(1)} + y^{(1)} \in \mathcal{F}_1$. Then $(x^{(1)}, 0) + (y^{(1)}, 0) \in \mathcal{F}$. But $(x^{(1)}, 0), (y^{(1)}, 0) \in \mathcal{K}$ (as $0 \in \mathcal{K}_2$), so $(x^{(1)}, 0), (y^{(1)}, 0) \in \mathcal{F}$. Hence $x^{(1)}, y^{(1)} \in \mathcal{F}_1$. This shows that $\mathcal{F}_1 \trianglelefteq \mathcal{K}_1$. Similarly, $\mathcal{F}_2 \trianglelefteq \mathcal{K}_2$. Next, we need to show that $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$. If $x^{(1)} \in \mathcal{F}_1$ and $x^{(2)} \in \mathcal{F}_2$, then $(x^{(1)}, 0), (0, x^{(2)}) \in \mathcal{F}$, so $(x^{(1)}, x^{(2)}) = (x^{(1)}, 0) + (0, x^{(2)}) \in \mathcal{F}$, i.e., $\mathcal{F}_1 \times \mathcal{F}_2 \subseteq \mathcal{F}$. Conversely, if $(x^{(1)}, x^{(2)}) \in \mathcal{F} \subseteq \mathcal{K}_1 \times \mathcal{K}_2$, then $(x^{(1)}, 0), (0, x^{(2)}) \in \mathcal{K}_1 \times \mathcal{K}_2$. Hence $(x^{(1)}, x^{(2)}) = (x^{(1)}, 0) + (0, x^{(2)})$ implies that $(x^{(1)}, 0), (0, x^{(2)}) \in \mathcal{F}$, so $x^{(1)} \in \mathcal{F}_1$ and $x^{(2)} \in \mathcal{F}_2$. This shows that $\mathcal{F} \subseteq \mathcal{F}_1 \times \mathcal{F}_2$. Therefore $\mathcal{F} \trianglelefteq \mathcal{K}_1 \times \mathcal{K}_2$ implies that $\mathcal{F}$ is a Cartesian product of some faces of $\mathcal{K}_1$ and $\mathcal{K}_2$.

Conversely, suppose that $\mathcal{F}_1 \trianglelefteq \mathcal{K}_1$ and $\mathcal{F}_2 \trianglelefteq \mathcal{K}_2$. Let $(x^{(1)}, x^{(2)}), (y^{(1)}, y^{(2)}) \in \mathcal{K}_1 \times \mathcal{K}_2$ satisfy $(x^{(1)}, x^{(2)}) + (y^{(1)}, y^{(2)}) \in \mathcal{F}_1 \times \mathcal{F}_2$. Then

$$x^{(1)}, y^{(1)} \in \mathcal{K}_1,\ x^{(1)} + y^{(1)} \in \mathcal{F}_1 \implies x^{(1)}, y^{(1)} \in \mathcal{F}_1,$$

and similarly, $x^{(2)}, y^{(2)} \in \mathcal{F}_2$. Hence $(x^{(1)}, x^{(2)}), (y^{(1)}, y^{(2)}) \in \mathcal{F}_1 \times \mathcal{F}_2$. This shows that $\mathcal{F}_1 \times \mathcal{F}_2$ is indeed a face of $\mathcal{K}_1 \times \mathcal{K}_2$. □

Of interest is also the minimal faces of convex sets contained in a Cartesian product of convex cones. By Proposition 2.2.19, we know that such minimal faces are Cartesian products of faces in the smaller cones. In fact, they are minimal faces of the projection of the convex sets.

Proposition 2.2.20. *Let $\mathcal{K}_i \subseteq \mathbb{V}_i$ be a convex cone for $i = 1, 2$, and let $\varnothing \neq \mathcal{S} \in \mathcal{K}_1 \times \mathcal{K}_2$ be a convex set. Define*

$$\mathcal{S}_1 := \left\{ \tilde{x}^{(1)} \in \mathbb{V}_1 : (\tilde{x}^{(1)}, x^{(2)}) \in \mathcal{S} \text{ for some } x^{(2)} \in \mathbb{V}_2 \right\},$$
$$\mathcal{S}_2 := \left\{ \tilde{x}^{(2)} \in \mathbb{V}_2 : (x^{(1)}, \tilde{x}^{(2)}) \in \mathcal{S} \text{ for some } x^{(1)} \in \mathbb{V}_1 \right\}.$$

Then $\mathcal{S}_1$ and $\mathcal{S}_2$ are convex sets, and

$$\operatorname{face}(\mathcal{S}, \mathcal{K}_1 \times \mathcal{K}_2) = \mathcal{F}_1 \times \mathcal{F}_2, \quad \text{where } \mathcal{F}_j = \operatorname{face}(\mathcal{S}_j, \mathcal{K}_j) \text{ for } j = 1, 2.$$

Proof. That $\mathcal{S}_1$ and $\mathcal{S}_2$ are convex sets is immediate.[4] By Proposition 2.2.19, $\operatorname{face}(\mathcal{S}, \mathcal{K}_1 \times \mathcal{K}_2) = \mathcal{F}_1 \times \mathcal{F}_2$ for some $\mathcal{F}_j \trianglelefteq \mathcal{K}_j$, $j = 1, 2$. Then $\mathcal{S}_1 \subseteq \mathcal{F}_1$[5] implies that $\operatorname{face}(\mathcal{S}_1, \mathcal{K}_1) \subseteq \mathcal{F}_1$. We show that $\mathcal{F}_1 = \operatorname{face}(\mathcal{S}_1, \mathcal{K}_1)$. By Proposition 2.2.5,

$$\varnothing \neq \mathcal{S} \cap \operatorname{ri}(\mathcal{F}_1 \times \mathcal{F}_2) = \mathcal{S} \cap (\operatorname{ri}(\mathcal{F}_1) \times \operatorname{ri}(\mathcal{F}_2)),$$

implying that there exists $(x^{(1)}, x^{(2)}) \in \mathcal{S}$ such that $x^{(j)} \in \mathcal{S}_j \cap \operatorname{ri}(\mathcal{F}_j)$ for $j = 1, 2$. Hence $\mathcal{S}_j \cap \operatorname{ri}(\mathcal{F}_j) \neq \varnothing$, i.e., $\mathcal{F}_j = \operatorname{face}(\mathcal{S}_j, \mathcal{K}_j)$ for $j = 1, 2$ by Proposition 2.2.5. □

[4] Let $\tilde{x}^{(1)}, \tilde{y}^{(1)} \in \mathcal{S}_1$ and $\alpha \in [0, 1]$. Then there exist $x^{(2)}, y^{(2)}$ such that $(\tilde{x}^{(1)}, x^{(2)}), (\tilde{y}^{(1)}, y^{(2)}) \in \mathcal{S}$. By convexity, $(\alpha\tilde{x}^{(1)} + (1-\alpha)\tilde{y}^{(1)}, \alpha x^{(2)} + (1-\alpha)y^{(2)} \in \mathcal{S}$, so $\alpha\tilde{x}^{(1)} + (1-\alpha)\tilde{y}^{(1)} \in \mathcal{S}_1$.

[5] For any $\tilde{x}^{(1)} \in \mathcal{S}_1$, $(\tilde{x}^{(1)}, x^{(2)}) \in \mathcal{S} \subseteq \mathcal{F}$, so $\tilde{x}^{(1)} \in \mathcal{F}_1$.

Chapter 3

Preliminaries on conic programming

In this chapter, we introduce some basic properties of conic programs relevant in this thesis, and occasionally discuss semidefinite programs as a special case. More comprehensive studies on semidefinite programming can be found in e.g., [5, 12, 22, 88, 90, 98].

A *(linear) conic program* is an optimization problem of the form

$$v_{\mathrm{P_{conic}}} = \sup_y \left\{ b^\top y : C - \mathcal{A}^* y \in \mathcal{K} \right\}, \tag{$\mathrm{P_{conic}}$}$$

where $(\mathbb{V}, \langle \cdot, \cdot \rangle_{\mathbb{V}})$ is a finite dimensional inner product space, $\mathcal{A} : \mathbb{V} \to \mathbb{R}^m$ is a linear map, $b \in \mathbb{R}^m$, $C \in \mathbb{V}$, and $\{0\} \neq \mathcal{K} \subseteq \mathbb{V}$ is a nonempty closed convex cone.

The conic program ($\mathrm{P_{conic}}$) is said to be *feasible* if there exists $y \in \mathbb{R}^m$ such that $C - \mathcal{A}^* y \in \mathcal{K}$ and *infeasible* if no such y exists. Any $y \in \mathbb{R}^m$ satisfying $Z := C - \mathcal{A}^* y \in \mathcal{K}$ is called a *feasible point* of ($\mathrm{P_{conic}}$), and Z is called a *feasible slack* of ($\mathrm{P_{conic}}$). More specifically, ($\mathrm{P_{conic}}$) is said to be *asymptotically feasible* if there exist sequences $\{Z^{(k)}\}_k \subset \mathcal{K}$ and $\{y^{(k)}\}_k \subset \mathbb{R}^m$ such that $Z^{(k)} + \mathcal{A}^* y^{(k)} \to C$ as $k \to \infty$. (Observe that if ($\mathrm{P_{conic}}$) is feasible, then ($\mathrm{P_{conic}}$) is asymptotically feasible.) Furthermore, ($\mathrm{P_{conic}}$) is said to be *weakly infeasible* if ($\mathrm{P_{conic}}$) is infeasible but asymptotically feasible, and *strongly infeasible* if ($\mathrm{P_{conic}}$) is not asymptotically feasible. We will further discuss asymptotic feasibility in Section 7.1.

We say ($\mathrm{P_{conic}}$) is *unbounded* if $v_{\mathrm{P_{conic}}} = +\infty$, i.e., there exists a sequence $\left\{y^{(k)}\right\}_k \subset \mathbb{R}^m$ such that $C - \mathcal{A}^* y^{(k)} \in \mathcal{K}$ for all k and $\lim_k b^\top y^{(k)} = +\infty$. We take $v_{\mathrm{P_{conic}}} = -\infty$ if ($\mathrm{P_{conic}}$) is infeasible. We say that ($\mathrm{P_{conic}}$) is *solvable* if its optimal value $v_{\mathrm{P_{conic}}}$ is *attained*, i.e., if there exists $y \in \mathbb{R}^m$ such that $C - \mathcal{A}^* y \in \mathcal{K}$ and $b^\top y = v_{\mathrm{P_{conic}}}$.

If $b \notin \mathrm{range}(\mathcal{A})$, then ($\mathrm{P_{conic}}$) is unbounded whenever ($\mathrm{P_{conic}}$) is feasible. To rule out this scenario, we impose the following assumption when studying the conic program ($\mathrm{P_{conic}}$):

Assumption 3.1. *There exists some $\breve{X} \in \mathcal{K}^*$ such that $\mathcal{A}(\breve{X}) = b$.*

Since replacing $\mathcal{A}$ by an onto linear map $\tilde{\mathcal{A}}$ with $\text{range}(\mathcal{A}^*) = \text{range}(\tilde{\mathcal{A}}^*)$ does not change the optimal value of (P_{conic}), we assume without loss of generality that $\mathcal{A}$ itself is onto:

Assumption 3.2. *The linear map $\mathcal{A}$ is onto.*

Under Assumption 3.2, $\mathcal{A}^* y \neq 0$ for all $y \neq 0$, and

$$\sigma_{\min}(\mathcal{A}^*) := \min_{\|y\|=1} \|\mathcal{A}^* y\|_{\mathbb{V}} > 0. \tag{3.1}$$

The optimization problem

$$v_{\text{D}_{\text{conic}}} = \inf_X \{\langle C, X\rangle_{\mathbb{V}} : \mathcal{A}(X) = b,\ X \in \mathcal{K}^*\} \tag{D_{conic}}$$

is often associated with the general conic program (P_{conic}), and is called the *dual* of (P_{conic}). We will explain in Section 3.2 the relationship between this dual program and (P_{conic}). We typically call the two programs (P_{conic})-(D_{conic}) a *primal-dual pair*.

We define the following notation for the feasible regions of (P_{conic}) and (D_{conic}):

$$\begin{aligned}
\mathcal{F}^y_{\text{P}_{\text{conic}}} &:= \{y \in \mathbb{R}^m : C - \mathcal{A}^* y \in \mathcal{K}\}, \\
\mathcal{F}^Z_{\text{P}_{\text{conic}}} &:= \{Z \in \mathbb{V} : Z = C - \mathcal{A}^* y \in \mathcal{K} \text{ for some } y \in \mathbb{R}^m\}, \\
\mathcal{F}_{\text{D}_{\text{conic}}} &:= \{X \in \mathbb{V} : \mathcal{A}(X) = b,\ X \in \mathcal{K}^*\}.
\end{aligned}$$

Having introduced the basic notation, we outline the organization of this chapter. We first introduce some important classes of conic programs that are often seen in practice in Section 3.1. In Section 3.2, we review the duality theory for (P_{conic}), including the strong duality theorem (Theorem 3.3.3). In Section 3.3, we introduce the Slater condition, also known as the strict feasibility, which is commonly assumed to hold for both (P_{conic}) and (D_{conic}) to ensure desirable properties. We discuss the consequences and equivalent conditions of the Slater condition.

3.1 Important classes of conic programs

The conic program (P_{conic}) generalizes several classes of important optimization problems:

- linear programs (LP), with $\mathbb{V} = \mathbb{R}^n$ and $\mathcal{K} = \mathbb{R}^n_+$;
- second order cone programs (SOCP), with $\mathbb{V} = \mathbb{R}^n$ and $\mathcal{K} = \mathcal{Q}^{n_1} \times \mathcal{Q}^{n_2} \times \cdots \times \mathcal{Q}^{n_k}$, where the positive integers $n_1, n_2, \ldots, n_k$ sum to n;

- semidefinite programs (SDP), with $\mathbb{V} = \mathbb{S}^n$ and $\mathcal{K} = \mathbb{S}^n_+$;
- copositive programs (CoP). with $\mathbb{V} = \mathbb{S}^n$ and $\mathcal{K} = \mathcal{C}^n$; and
- mixed conic programs.

In the following we introduce each of these classes of problems, with emphasis on SDP.

Linear programs

A *linear program* (LP) is an optimization problem over the nonnegative orthant. A primal-dual pair of LP is typically of the form

$$v_{\mathrm{P_{LP}}} := \max_y \left\{ b^\top y : c - A^\top y \geq 0 \right\}, \tag{$\mathrm{P_{LP}}$}$$

$$v_{\mathrm{D_{LP}}} := \min_x \left\{ c^\top x : Ax = b,\ x \geq 0 \right\}, \tag{$\mathrm{D_{LP}}$}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$.

Linear programs enjoy a very important property: as long as the optimal value $v_{\mathrm{P_{LP}}}$ is finite, it is attained and equals to the dual optimal value $v_{\mathrm{D_{LP}}}$, which is also attained. We state this classical result, which can be found in many textbooks, e.g., [93, Theorem 3.4, Theorem 5.2].

Theorem 3.1.1. *The linear program* ($\mathrm{P_{LP}}$) *is either infeasible (i.e., there exists no y satisfying $c - A^\top y \geq 0$), or unbounded (i.e., $v_{\mathrm{P_{LP}}} = +\infty$), or solvable (i.e., $v_{\mathrm{P_{LP}}} = b^\top \bar{y}$ for some $\bar{y} \in \mathbb{R}^m$ satisfying $c - A^\top \bar{y} \geq 0$). If* ($\mathrm{P_{LP}}$) *is solvable, then its dual program* ($\mathrm{D_{LP}}$) *is also solvable and $v_{\mathrm{P_{LP}}} = v_{\mathrm{D_{LP}}}$.*

The first part of Theorem 3.1.1 is part of what is commonly called the *fundamental theorem of linear programming*, and the second part of Theorem 3.1.1 is commonly known as the strong duality theorem of linear programming.

Second order cone programs

A *second order cone program* (SOCP) is an optimization problem over $\mathcal{K} = \mathcal{Q}^{n_1} \times \mathcal{Q}^{n_2} \times \cdots \mathcal{Q}^{n_k}$ of second order cones. (Note that $\mathcal{K}$ is self-dual.) A primal-dual pair of SOCP is typically of the form

$$v_{\mathrm{P_{SOCP}}} := \sup_y \left\{ b^\top y : z = c - A^\top y \in \mathcal{K} \right\}, \tag{$\mathrm{P_{SOCP}}$}$$

$$v_{\mathrm{D_{SOCP}}} := \inf_x \left\{ c^\top x : Ax = b,\ x \in \mathcal{K} \right\}, \tag{$\mathrm{D_{SOCP}}$}$$

where $A = [A_1, A_2, \ldots, A_k] \in \mathbb{R}^{m\times(n_1+n_2+\ldots+n_k)}$, $b \in \mathbb{R}^m$ and $c = (c_1; c_2; \ldots; c_k) \in \mathbb{R}^{n_1+n_2+\ldots+n_k}$.

(See, e.g., [4] for further details on SOCP.)

Semidefinite programs

A *semidefinite program* (SDP) is an optimization problem over the cone of positive semidefinite matrices. LP and SOCP are both special cases of SDP. A primal-dual pair of (linear) SDP is typically of the form

$$v_{\mathrm{P}} := \sup_y \left\{ b^\top y : C - \mathcal{A}^* y \succeq 0 \right\}, \tag{P}$$

$$v_{\mathrm{D}} := \inf_X \left\{ \langle C, X\rangle : \mathcal{A}(X) = b,\ X \succeq 0 \right\}, \tag{D}$$

where

$$\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m : X \mapsto \begin{pmatrix} \langle A_1, X\rangle \\ \vdots \\ \langle A_m, X\rangle \end{pmatrix}, \quad A_1, \ldots, A_m \in \mathbb{S}^n$$

is a linear map, $\mathcal{A}^* y := \sum_{j=1}^m y_j A_j$ is the adjoint of $\mathcal{A}$ (with respect to the inner product $\langle \cdot, \cdot \rangle$), $b \in \mathbb{R}^m$ and $C \in \mathbb{S}^n$.

Under Assumption 3.1, $X \in \mathbb{S}^n$ is feasible for (D) if and only if $X = \hat{X} + \mathcal{V}^* v \succeq 0$ for some $v \in \mathbb{R}^s$, where $\mathcal{V} : \mathbb{S}^n \to \mathbb{R}^s$ is a linear map with $\mathrm{range}(\mathcal{V}^*) = \ker(\mathcal{A})$. Hence (D) equals

$$\langle C, \hat{X}\rangle + \inf_v \left\{ (\mathcal{V}(C))^\top v : \hat{X} + \mathcal{V}^* v \succeq 0 \right\},$$

which is in the same form as (P).

We define the following notation for the feasible regions of (P) and (D):

$$\begin{aligned} \mathcal{F}_{\mathrm{P}}^y &:= \{ y \in \mathbb{R}^m : C - \mathcal{A}^* y \succeq 0 \}, \\ \mathcal{F}_{\mathrm{P}}^Z &:= \{ Z \in \mathbb{S}^n : Z = C - \mathcal{A}^* y \succeq 0 \text{ for some } y \in \mathbb{R}^m \}, \\ \mathcal{F}_{\mathrm{D}} &:= \{ X \in \mathbb{S}^n : \mathcal{A}(X) = b,\ X \succeq 0 \}. \end{aligned}$$

Copositive programs

A *copositive program* (CoP) is an optimization problem over the copositive cone $\mathcal{C}^n$. A primal-dual pair of CoP is typically of the form

$$v_{\mathrm{P_{CoP}}} := \sup_y \left\{ b^\top y : C - \mathcal{A}^* y \in \mathcal{C}^n \right\}, \tag{$\mathrm{P_{CoP}}$}$$

$$v_{\mathrm{D_{CoP}}} := \inf_X \left\{ \langle C, X\rangle : \mathcal{A}(X) = b,\ X \in (\mathcal{C}^n)^* \right\}, \tag{$\mathrm{D_{CoP}}$}$$

where $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ is a linear map, $b \in \mathbb{R}^m$ and $C \in \mathbb{S}^n$. Observe that since $\mathcal{C}^n$ is not self-dual, the dual program ($\mathrm{D_{CoP}}$) optimizes over the completely positive cone, which is different from the copositive cone.

(See, e.g., [33] and [5, Chapter 8] for further details on copositive programming.)

Mixed conic programs

In practice, the inner product space $\mathbb{V}$ and the cone $\mathcal{K}$ are often Cartesian products, i.e., $\mathbb{V} = \mathbb{V}_1 \times \cdots \times \mathbb{V}_k$ and $\mathcal{K} = \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_k$, where $\mathbb{V}_j$ is an inner product space and $\mathcal{K}_j \subseteq \mathbb{V}_j$ is a nonempty closed convex cone; and often $\mathcal{K}_i$ is one of $\mathbb{R}^{n_i}$, $\mathcal{Q}^{n_i}$ or $\mathbb{S}^{n_i}_+$. A typical primal-dual pair of mixed conic programs is of the form

$$\begin{aligned} v_{\mathrm{P_m}} := \quad & \sup_y \left\{ b^\top y : C - \mathcal{A}^* y \in \mathcal{K}_1 \times \mathcal{K}_2 \times \cdots \times \mathcal{K}_k \right\} && (\mathrm{P_m}) \\ = \quad & \sup_{y, Z^{(1)}, \ldots, Z^{(k)}} \left\{ b^\top y : Z^{(j)} = C^{(j)} - (\mathcal{A}^{(j)})^* y \in \mathcal{K}_j,\ \forall\, j \in 1:k \right\}, \\ v_{\mathrm{D_m}} := \quad & \inf_X \{ \langle C, X \rangle_{\mathbb{V}} : \mathcal{A}(X) = b,\ X \in \mathcal{K}_1^* \times \mathcal{K}_2^* \times \cdots \times \mathcal{K}_k^* \} && (\mathrm{D_m}) \\ = \quad & \inf_{X^{(1)}, \ldots, X^{(k)}} \left\{ \sum_{j=1}^k \langle C^{(j)}, X^{(j)} \rangle_{\mathbb{V}_j} : \sum_{j=1}^k \mathcal{A}^{(j)}(X^{(j)}) = b,\ X^{(j)} \in \mathcal{K}_j^*,\ \forall\, j = 1:k \right\}, \end{aligned}$$

where

$$\mathcal{A} \ : \ \mathbb{V}_1 \times \cdots \times \mathbb{V}_k \to \mathbb{R}^m \ : \ (X^{(1)}, \ldots, X^{(k)}) \mapsto \sum_{j=1}^k \mathcal{A}^{(j)}(X^{(j)}),$$

$\mathcal{A}^{(j)} : \mathbb{V}_j \to \mathbb{R}^m$ is a linear map, $b \in \mathbb{R}^m$, and $C = (C^{(1)}, \ldots, C^{(k)}) \in \mathbb{V}_1 \times \cdots \times \mathbb{V}_k$.

3.2 Duality theory

A common strategy for estimating $v_{\mathrm{P_{conic}}}$ is to get an upper bound via the Lagrangian. The *Lagrangian* of ($\mathrm{P_{conic}}$) is a function $\mathrm{L} : \mathbb{R}^m \times \mathbb{V} \to \mathbb{R}$ defined by

$$\mathrm{L}(y, X) = b^\top y + \langle X, C - \mathcal{A}^* y \rangle_{\mathbb{V}} = (b - \mathcal{A}(X))^\top y + \langle C, X \rangle_{\mathbb{V}}.$$

Now note that for all $X \in \mathcal{K}^*$,

$$v_{\mathrm{P_{conic}}} \leq \sup_y \mathrm{L}(y, X) = \begin{cases} \langle C, X \rangle_{\mathbb{V}} & \text{if } \mathcal{A}(X) = b, \\ +\infty & \text{otherwise.} \end{cases}$$

Therefore we get that $v_{\mathrm{P_{conic}}} \leq v_{\mathrm{D_{conic}}}$, where

$$v_{\mathrm{D_{conic}}} = \inf_X \{\langle C, X\rangle_{\mathbb{V}} : \mathcal{A}(X) = b,\ X \in \mathcal{K}^*\}$$

is called the *Lagrangian dual* of ($\mathrm{P_{conic}}$).

In the following sections we outline some important notions associated with conic programs: (1) subspace form, (2) weak and strong duality, and (3) strict complementarity. Blekherman *et al.* [12] gives a very good list of issues concerning general conic programs.

3.2.1 Subspace form

We first discuss the subspace form for SDP. The primal-dual pair of conic programs ($\mathrm{P_{conic}}$)-($\mathrm{D_{conic}}$) are often expressed in terms of the explicit algebraic description of the linear subspace $\mathcal{L} := \mathrm{range}(\mathcal{A}^*)$; for instance, in SDP, we use the matrices $A_1, \ldots, A_m$. It is often useful to rewrite ($\mathrm{P_{conic}}$)-($\mathrm{D_{conic}}$) in *subspace form*, which is independent of the algebraic description, when studying the theoretical aspects of the conic programs.

Let $(\tilde{X}, \tilde{y}, \tilde{Z})$ satisfy the linear equations $\mathcal{A}(\tilde{X}) = b$, $C - \mathcal{A}^*\tilde{y} = \tilde{Z}$. Then

$$\begin{aligned}
\sup_y \left\{ b^\top y : C - \mathcal{A}^* y \in \mathcal{K} \right\} &= \sup_y \left\{ \langle \tilde{X}, \mathcal{A}^* y\rangle_{\mathbb{V}} : C - \mathcal{A}^* y \in \mathcal{K} \right\} \\
&= \sup_{y,Z} \left\{ \langle \tilde{X}, C - Z\rangle_{\mathbb{V}} : Z = C - \mathcal{A}^* y \in \mathcal{K} \right\} \\
&= \langle \tilde{X}, C\rangle_{\mathbb{V}} - \inf_Z \left\{ \langle \tilde{X}, Z\rangle_{\mathbb{V}} : Z \in (C + \mathrm{range}(\mathcal{A}^*)) \cap \mathcal{K} \right\},
\end{aligned}$$

and

$$\begin{aligned}
\inf_X \{\langle C, X\rangle_{\mathbb{V}} : \mathcal{A}(X) = b,\ X \in \mathcal{K}^*\} &= \inf_X \left\{ \langle C, X\rangle_{\mathbb{V}} : X \in (\tilde{X} + \ker(\mathcal{A})) \cap \mathcal{K}^* \right\} \\
&= \inf_X \left\{ \langle \tilde{Z} + \mathcal{A}^*\tilde{y}, X\rangle_{\mathbb{V}} : X \in (\tilde{X} + \ker(\mathcal{A})) \cap \mathcal{K}^* \right\} \\
&= b^\top \tilde{y} + \inf_X \left\{ \langle \tilde{Z}, X\rangle_{\mathbb{V}} : X \in (\tilde{X} + \ker(\mathcal{A})) \cap \mathcal{K}^* \right\}.
\end{aligned}$$

Using the fact that $C + \mathcal{L} = \tilde{Z} + \mathcal{L}$, we have

$$v_{\mathrm{P_{conic}}} = \langle \tilde{X}, C\rangle_{\mathbb{V}} - \inf_Z \left\{ \langle \tilde{X}, Z\rangle_{\mathbb{V}} : Z \in (\tilde{Z} + \mathcal{L}) \cap \mathcal{K} \right\}, \tag{3.2a}$$

$$v_{\mathrm{D_{conic}}} = \quad b^\top \tilde{y} \quad + \inf_X \left\{ \langle \tilde{Z}, X\rangle_{\mathbb{V}} : X \in (\tilde{X} + \mathcal{L}^\perp) \cap \mathcal{K}^* \right\}. \tag{3.2b}$$

We call the symmetric primal-dual pair (3.2a)-(3.2b) the *subspace form* for ($\mathrm{P_{conic}}$)-($\mathrm{D_{conic}}$). In the literature (e.g. [62, 68, 86, 91]...) the subspace form is often used in place of ($\mathrm{P_{conic}}$)-($\mathrm{D_{conic}}$) in the study of the properties of the feasible regions and their relationship to strong duality,

because those properties are often intrinsic, i.e., independent of algebraic expression of the linear map $\mathcal{A}$ and the subspace form liberates those properties from being dependent on a particular choice of algebraic description.

3.2.2 Weak and strong duality

As a result of the derivation of the Lagrangian dual ($\mathrm{D}_{\mathrm{conic}}$), we obtain the *weak duality theorem*:

Theorem 3.2.1 (Weak duality theorem)**.** *If y is feasible for ($\mathrm{P}_{\mathrm{conic}}$) and X is feasible for ($\mathrm{D}_{\mathrm{conic}}$), then $b^\top y \leq \langle C, X\rangle_{\mathbb{V}}$. In particular, $v_{\mathrm{P}_{\mathrm{conic}}} \leq v_{\mathrm{D}_{\mathrm{conic}}}$.*

The discrepancy between the primal and dual optimal value, i.e., the difference $v_{\mathrm{D}_{\mathrm{conic}}} - v_{\mathrm{P}_{\mathrm{conic}}}$, is called the *duality gap*. More generally, given a primal feasible solution (y, Z) of ($\mathrm{P}_{\mathrm{conic}}$) and a dual feasible solution X of ($\mathrm{D}_{\mathrm{conic}}$), we define their duality gap as the difference in objective values:

$$\text{duality gap between } (y, Z) \text{ and } X \;:=\; \langle C, X\rangle_{\mathbb{V}} - b^\top y \;=\; \langle X, Z\rangle_{\mathbb{V}}.$$

By the weak duality theorem, the duality gap is always nonnegative. Since the dual program ($\mathrm{D}_{\mathrm{conic}}$) is set up to provide an upper bound on the primal optimal value $v_{\mathrm{P}_{\mathrm{conic}}}$, intuitively ($\mathrm{D}_{\mathrm{conic}}$) is most useful when its optimal value $v_{\mathrm{D}_{\mathrm{conic}}}$ equals to $v_{\mathrm{P}_{\mathrm{conic}}}$, i.e., there is a zero duality gap. *Strong duality* is said to hold for ($\mathrm{P}_{\mathrm{conic}}$) if $v_{\mathrm{P}_{\mathrm{conic}}} = v_{\mathrm{D}_{\mathrm{conic}}}$ and ($\mathrm{D}_{\mathrm{conic}}$) has an optimal solution. If strong duality holds for ($\mathrm{P}_{\mathrm{conic}}$), then $X \in \mathbb{V}$ is an optimal solution if and only if there exist $y \in \mathbb{R}^m$ and $Z \in \mathbb{V}$ such that the following system holds:

$$\begin{array}{rll} Z = C - \mathcal{A}^* y, & Z \in \mathcal{K}, & \text{(dual feasibility)} \\ \mathcal{A}(X) = b, & X \in \mathcal{K}^*, & \text{(primal feasibility)} \\ \langle X, Z\rangle_{\mathbb{V}} = 0. & & \text{(complementary slackness)} \end{array} \tag{3.3}$$

While (3.3) is sufficient (but not always necessary) for a primal feasible solution (y, Z) and dual feasible solution X to be optimal, respectively, for ($\mathrm{P}_{\mathrm{conic}}$) and ($\mathrm{D}_{\mathrm{conic}}$), the necessity of (3.3) does hold for a primal-dual pair of optimal solutions where there is a zero duality gap between them.

Proposition 3.2.2. *Let $(X, y, Z) \in \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n$. Then the following are equivalent.*

(1) (X, y, Z) satisfies (3.3)*.*

(2) X is optimal for ($\mathrm{P}_{\mathrm{conic}}$), (y, Z) is optimal for ($\mathrm{D}_{\mathrm{conic}}$) and $v_{\mathrm{P}_{\mathrm{conic}}} = v_{\mathrm{D}_{\mathrm{conic}}}$.

Proof. If Item (1) holds, then (y, Z) is feasible for $(\mathrm{P}_{\mathrm{conic}})$ and X is feasible for $(\mathrm{D}_{\mathrm{conic}})$. Moreover, the duality gap is zero (because $\langle C, X\rangle_{\mathbb{V}} - b^\top y = \langle X, Z\rangle_{\mathbb{V}} = 0$), so by the weak duality theorem (Theorem 3.2.1), we have $v_{\mathrm{P}_{\mathrm{conic}}} = b^\top y = \langle C, X\rangle_{\mathbb{V}} = v_{\mathrm{D}_{\mathrm{conic}}}$. Hence Item (2) holds.

Conversely, if Item (2) holds, then $\langle X, Z\rangle_{\mathbb{V}} = v_{\mathrm{P}_{\mathrm{conic}}} - v_{\mathrm{D}_{\mathrm{conic}}} = 0$. Hence Item (1) holds. □

A common sufficient condition for strong duality is the Slater condition together with finite optimal value, see Section 3.3 below.

3.2.3 Strict complementarity

In this section, we are concerned with the optimal solutions of $(\mathrm{P}_{\mathrm{conic}})$ and $(\mathrm{D}_{\mathrm{conic}})$ when there is a zero duality gap.

Define

$$\begin{aligned}
\mathcal{O}_{\mathrm{conic}} &:= \{(X, y, Z) \in \mathbb{V} \times \mathbb{R}^m \times \mathbb{V} : (X, y, Z) \text{ satisfies (3.3)}\}, \\
\mathcal{O}^y_{\mathrm{P}_{\mathrm{conic}}} &:= \{y \in \mathbb{R}^m : (X, y, Z) \in \mathcal{O}_{\mathrm{conic}} \text{ for some } X, Z\}, \\
\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}} &:= \{Z \in \mathbb{V} : (X, y, Z) \in \mathcal{O}_{\mathrm{conic}} \text{ for some } X, y\}, \\
\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}} &:= \{X \in \mathbb{V} : (X, y, Z) \in \mathcal{O}_{\mathrm{conic}} \text{ for some } y, Z\}
\end{aligned}$$

to be the sets of primal-dual solutions of $(\mathrm{P}_{\mathrm{conic}})$-$(\mathrm{D}_{\mathrm{conic}})$ *with zero duality gap*. It is immediate that

$$\begin{aligned}
\mathcal{O}_{\mathrm{conic}} = \{(X, y, Z) \in \mathbb{V} \times \mathbb{R}^m \times \mathbb{V} : \quad & Z = C - \mathcal{A}^* y, \quad Z \in \mathcal{K}, \\
& \mathcal{A}(X) = b, \quad X \in \mathcal{K}^*, \\
& \langle C, X\rangle_{\mathbb{V}} - b^\top y = 0 \ \Big\} \\
= \Big\{(X, y, Z) : Z = C - \mathcal{A}^* y,\ \mathcal{A}(X) = b,\ \langle C, X\rangle_{\mathbb{V}} - b^\top y = 0\Big\} & \cap (\mathcal{K}^* \times \mathbb{R}^m \times \mathcal{K})
\end{aligned}$$

is the intersection of an affine subspace and a closed convex cone, so $\mathcal{O}_{\mathrm{conic}}$ is a convex set. As mentioned in the previous section, (y, Z) being optimal for $(\mathrm{P}_{\mathrm{conic}})$ and X being optimal for $(\mathrm{D}_{\mathrm{conic}})$ do not imply that $(X, y, Z) \in \mathcal{O}_{\mathrm{conic}}$ since there may be a nonzero duality gap. Note that any $Z \in \mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}$ and $X \in \mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}$ are *complementary*, i.e., $\langle X, Z\rangle_{\mathbb{V}} = 0$.

We first define maximally complementary solutions.

Definition 3.2.3. *Feasible primal-dual solutions $\bar{Z} \in \mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}$ and $\bar{X} \in \mathcal{F}_{\mathrm{D}_{\mathrm{conic}}}$ are maximally complementary if $\bar{Z} \in \mathrm{ri}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}})$ and $\bar{X} \in \mathrm{ri}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}})$.*

Note that any maximally complementary solutions $(\bar{Z}, \bar{X})$ not only are optimal for ($\mathrm{P}_{\mathrm{conic}}$)-($\mathrm{D}_{\mathrm{conic}}$), but also have a zero duality gap. Moreover, by Proposition 2.2.5, $\bar{Z} \in \mathrm{ri}(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K}))$ and $\bar{X} \in \mathrm{ri}(\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*))$. The complementarity of $\bar{Z}$ and $\bar{X}$ implies that $\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})$ and $\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*)$ forms a *complementarity partition* of $\mathcal{K}, \mathcal{K}^*$ [91]:

Proposition 3.2.4. *Suppose that* $\mathcal{O}_{\mathrm{conic}} \neq \varnothing$. *Then* $\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*) \subseteq \left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c$ *and* $\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K}) \subseteq (\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*))^c$.

Proof. Let $\bar{Z} \in \mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}$ and $\bar{X} \in \mathcal{F}_{\mathrm{D}_{\mathrm{conic}}}$ be maximally complementary solutions. Then by Proposition 3.2.2, any $X \in \mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}$ must satisfy $\langle X, \bar{Z}\rangle_{\mathbb{V}} = 0$, so $\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}} \subseteq \mathcal{K}^* \cap \left\{\bar{Z}\right\}^{\perp}$. But $\bar{Z} \in \mathrm{ri}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}) \subseteq \mathrm{ri}\left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)$, so

$$\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}} \subseteq \mathcal{K}^* \cap \left\{\bar{Z}\right\}^{\perp} = \mathcal{K}^* \cap \mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})^{\perp} = \left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c,$$

and similarly

$$\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}} \subseteq (\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*))^c.$$

Hence $\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*) \subseteq \left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c$ and $\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K}) \subseteq (\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*))^c$. □

In general, however, $\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})$ may be a proper face of $\left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c$; in that case, ($\mathrm{P}_{\mathrm{conic}}$)-($\mathrm{D}_{\mathrm{conic}}$) would not have *strictly complementary* solutions, whose existence is often assumed in the convergence proofs of interior point methods.

Definition 3.2.5. *Feasible primal-dual solutions* $\bar{Z} \in \mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}$ *and* $\bar{X} \in \mathcal{F}_{\mathrm{D}_{\mathrm{conic}}}$ *are* strictly complementary *if they are maximally complementary and* $\bar{Z} \in \mathrm{ri}(\bar{X}^{\perp} \cap \mathcal{K})$ *or* $\bar{X} \in \mathrm{ri}(\bar{Z}^{\perp} \cap \mathcal{K}^*)$ *holds.*

The existence of strictly complementary solutions is equivalent to $\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K}), \mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*)$ being *strict complementary partition* of $\mathcal{K}, \mathcal{K}^*$:

Proposition 3.2.6. *Any maximally complementary solutions are strictly complementary solutions if and only if* $\left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c = \mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{F}^*)$ *or* $(\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*))^c = \mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{F})$ *holds.*

Proof. Suppose that $\left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c = \mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{F}^*)$, and let $\bar{Z} \in \mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}$ and $\bar{X} \in \mathcal{F}_{\mathrm{D}_{\mathrm{conic}}}$ be any maximally complementary solutions. Then $\bar{Z} \in \mathrm{ri}(\mathcal{O}_{\mathrm{P}_{\mathrm{conic}}})$ implies that $\left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c = \left\{\bar{Z}\right\}^{\perp} \cap \mathcal{K}^*$ and

$$\bar{X} \in \mathrm{ri}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}) \subseteq \mathrm{ri}\left(\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{F}^*)\right) = \mathrm{ri}\left(\left\{\bar{Z}\right\}^{\perp} \cap \mathcal{K}^*\right).$$

Similarly, $(\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*))^c = \mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{F})$ implies that $\bar{Z} \in \mathrm{ri}\left(\left\{\bar{X}\right\}^{\perp} \cap \mathcal{K}\right)$.

Conversely, suppose that $\bar{X} \in \mathrm{ri}\left(\{\bar{Z}\}^{\perp} \cap \mathcal{K}^*\right) = \mathrm{ri}\left(\left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c\right)$. Then $\bar{X} \in \mathcal{O}_{\mathrm{D}_{\mathrm{conic}}} \cap \mathrm{ri}\left(\left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c\right)$, implying that $\left(\mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^c = \mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*)$ by Proposition 2.2.5. Similarly $\bar{Z} \in \mathrm{ri}\left(\{\bar{X}\}^{\perp} \cap \mathcal{K}\right)$ implies that $(\mathrm{face}(\mathcal{O}_{\mathrm{D}_{\mathrm{conic}}}, \mathcal{K}^*))^c = \mathrm{face}(\mathcal{O}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{F})$. □

We close this section by stating a classical result on LP, that if an LP is solvable then it has strictly complementary solutions.

Theorem 3.2.7. *[46] If the linear program* ($\mathrm{P_{LP}}$) *and its dual* ($\mathrm{D_{LP}}$) *are both feasible, then* ($\mathrm{P_{LP}}$)-($\mathrm{D_{LP}}$) *has strictly complementary solutions, i.e., there exist an optimal solution* $\bar{y}$ *for* ($\mathrm{P_{LP}}$) *and an optimal solution* $\bar{x}$ *for* ($\mathrm{D_{LP}}$) *such that* $\bar{z} + \bar{x} := (c - A^\top \bar{y}) + \bar{x} > 0$.

3.3 Slater condition and minimal face

The *Slater condition* or *strict feasibility* on ($\mathrm{P_{conic}}$) is that

$$\exists\, \tilde{y} \in \mathbb{R}^m \;\; \text{s.t.} \;\; C - \mathcal{A}^*\tilde{y} \in \mathrm{ri}(\mathcal{K}),$$

and we call such a vector $\tilde{y}$ a *Slater point*. Similarly, the Slater condition is said to hold for ($\mathrm{D_{conic}}$) if there exists $\tilde{X} \in \mathrm{ri}(\mathcal{K}^*)$ such that $\mathcal{A}(\tilde{X}) = b$, and $\tilde{X}$ is called a *Slater point*.

The Slater condition is closely related to the notion of *well-posedness* of conic programs. If a conic program satisfies the Slater condition, then the conic program would remain feasible under any "small" perturbation on the input data. If a conic program fails the Slater condition, then the conic program is *ill-posed*, i.e., there exists arbitrarily small perturbation on the input data that results in a new infeasible conic program, and there also exists arbitrarily small perturbation on the input data that leaves the conic program feasible. We mention in passing that, for any given conic program, we can define its *distance to ill-posedness* (see e.g.,[41, 75, 76] and the references therein), which essentially quantifies how close the conic program is to failing the Slater condition.

Before we further explain the notion of the Slater condition, we first highlight its importance in Section 3.3.1. Then we explain its connection with the notion of the minimal face for conic programs in Section 3.3.2. A conic program that fails the Slater condition can be *regularized* by reducing the program to its minimal face; the minimal face can be found using a theorem of the alternative, see Theorem 3.3.10 in Section 3.3.3.

Some examples of semidefinite programs failing the Slater condition are given in Section 7.2.

3.3.1 Implications of the Slater condition

The Slater condition is useful for several reasons. It guarantees some desirable behavior of the SDP such as bounded dual sublevel sets and zero duality gap and is the basis for the notion of the central path, which is essential for the theory of interior point methods for solving SDP.

Below we describe several implications of Slater conditions, in conic programs and in SDP.

Compactness of dual sublevel sets

If the Slater condition holds for ($\mathrm{P}_{\mathrm{conic}}$) and if the cone $\mathcal{K}$ has nonempty interior (which is the case for all the conic programs introduced in Section 3.1), then the *dual sublevel sets*, defined in (3.4) below, are bounded. In particular, the set of dual optimal solutions is bounded, which is often an important property that guarantees the stability of interior point methods.

Proposition 3.3.1. *Suppose that* ($\mathrm{P}_{\mathrm{conic}}$) *satisfies the Slater condition and that* $\mathcal{K}$ *has nonempty interior. Then for any* $\alpha \in \mathbb{R}$*, the sublevel set*

$$\mathcal{S}_\alpha := \{X \in \mathbb{V} : \langle C, X\rangle_{\mathbb{V}} \leq \alpha,\ \mathcal{A}(X) = b,\ X \in \mathcal{K}^*\} \tag{3.4}$$

is compact.

Before we prove Proposition 3.3.1, we first prove a minor technical result.

Lemma 3.3.2. *Suppose that* $\mathcal{K} \subseteq \mathbb{V}$ *is a closed convex cone with nonempty interior and* $Z \in \mathrm{int}(\mathcal{K})$*. Then there exists a constant* $\delta > 0$ *such that for any* $X \in \mathcal{K}^*$*,* $\langle X, Z\rangle_{\mathbb{V}} \geq \delta\|X\|_{\mathbb{V}}$*.*

Proof. The optimization problem

$$\delta := \inf_X \{\langle Z, X\rangle_{\mathbb{V}} : X \in \mathcal{K}^*,\ \|X\|_{\mathbb{V}} = 1\} \tag{3.5}$$

has a compact feasible region and linear objective, so $\delta = \langle Z, \bar{X}\rangle_{\mathbb{V}} \geq 0$ for some $\bar{X} \in \mathcal{K}^*$ with $\|\bar{X}\|_{\mathbb{V}} = 1$. Suppose that $\delta = 0$. Pick a small $\epsilon > 0$ such that $Z - \epsilon\bar{X} \in \mathcal{K}$ (which exists because $Z \in \mathrm{int}(\mathcal{K})$). Then $0 \leq \langle Z - \epsilon\bar{X}, \bar{X}\rangle_{\mathbb{V}} = -\epsilon$, which is absurd. Hence we must have $\delta > 0$, and by (3.5) we get $\langle X, Z\rangle_{\mathbb{V}} \geq \delta\|X\|_{\mathbb{V}}$ for all $X \in \mathcal{K}^*$. □

Remark. When $\mathcal{K} = \mathbb{S}^n_+$, the constant δ equals the smallest eigenvalue of Z (which is positive since $Z \in \mathrm{int}(\mathbb{S}^n_+) = \mathbb{S}^n_{++}$).

Proof of Proposition 3.3.1. If $\alpha < v_{\mathrm{D}_{\mathrm{conic}}}$, then $\mathcal{S}_\alpha$ is empty, hence compact. Suppose that $\alpha \geq v_{\mathrm{D}_{\mathrm{conic}}}$. It is immediate that the set $\mathcal{S}_\alpha$ is closed. To see that it is compact, let $\tilde{Z} = C - \mathcal{A}^*\tilde{y} \in \mathrm{int}(\mathcal{K})$. Then by Lemma 3.3.2 there exists a constant $\delta > 0$ such that for any $X \in \mathcal{K}^*$, $\langle X, \tilde{Z}\rangle_{\mathbb{V}} \geq \delta\|X\|_{\mathbb{V}}$. Consequently, for any feasible solution X of ($\mathrm{P}_{\mathrm{conic}}$),

$$\delta\|X\|_{\mathbb{V}} \leq \langle X, \tilde{Z}\rangle = \langle C, X\rangle - b^\top\tilde{y} \leq \alpha - b^\top\tilde{y}.$$

Hence $\|X\|_{\mathbb{V}} \leq \frac{\alpha - b^\top\hat{y}}{\delta}$ for all $X \in \mathcal{S}_\alpha$, i.e., $\mathcal{S}_\alpha$ is bounded. □

Sufficiency for strong duality

A sufficient condition for strong duality of ($\mathrm{P}_{\mathrm{conic}}$) is the Slater condition on ($\mathrm{P}_{\mathrm{conic}}$) together with $v_{\mathrm{P}_{\mathrm{conic}}}$ being finite.

Theorem 3.3.3. *Consider the conic program* ($\mathrm{P}_{\mathrm{conic}}$) *and its dual* ($\mathrm{D}_{\mathrm{conic}}$). *Suppose that there exists* $\tilde{y} \in \mathbb{R}^m$ *such that* $C - \mathcal{A}^*\tilde{y} \in \mathrm{ri}(\mathcal{K})$, *and that* $v_{\mathrm{P}_{\mathrm{conic}}}$ *is finite. Then* $v_{\mathrm{P}_{\mathrm{conic}}} = v_{\mathrm{D}_{\mathrm{conic}}}$, *and* $v_{\mathrm{D}_{\mathrm{conic}}}$ *is attained.*

Proof. Let $\tilde{C} = C - \mathcal{A}^*\tilde{y}$. Then

$$v_{\mathrm{P}_{\mathrm{conic}}} = b^\top\tilde{y} + \sup_y \left\{ b^\top y : \tilde{C} - \mathcal{A}^*y \in \mathcal{K} \right\}, \tag{3.6a}$$

$$v_{\mathrm{D}_{\mathrm{conic}}} = b^\top\tilde{y} + \inf_X \left\{ \langle \tilde{C}, X\rangle_{\mathbb{V}} : \mathcal{A}(X) = b,\ X \in \mathcal{K}^* \right\}; \tag{3.6b}$$

also, ($\mathrm{P}_{\mathrm{conic}}$) and (3.6a) (resp., ($\mathrm{D}_{\mathrm{conic}}$) and (3.6b)) have the same feasible region and the same optimal solution set.

Note that $\tilde{C} \in \mathcal{K}$, so $\tilde{C} - \mathcal{A}^*y \in \mathcal{K}$ implies that $\mathcal{A}^*y \in \mathrm{span}(\mathcal{K})$. We first consider the trivial case that $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\mathcal{K}) = \{0\}$, then the case that $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\mathcal{K}) \neq \{0\}$.

- *Case 1.* $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\mathcal{K}) = \{0\}$.

  In this case, $y = 0$ is the only feasible solution, so $v_{\mathrm{P}_{\mathrm{conic}}} = b^\top\tilde{y}$.

  Since $v_{\mathrm{P}_{\mathrm{conic}}}$ is finite, we must have that $b = \mathcal{A}(\breve{X})$ for some $\breve{X} \in \mathbb{V}$. Since $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\mathcal{K}) = \{0\}$, we have $\ker(\mathcal{A}) + \mathrm{span}(\mathcal{K})^\perp = (\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\mathcal{K}))^\perp = \mathbb{V}$, implying that $\breve{X} = \breve{X}^{(1)} + \breve{X}^{(2)}$ for some $\breve{X}^{(1)} \in \ker(\mathcal{A})$ and $\breve{X}^{(2)} \in (\mathrm{span}(\mathcal{K}))^\perp$. Hence $b = \mathcal{A}(\breve{X}) = \mathcal{A}(\breve{X}^{(2)})$. Moreover, $\breve{X}^{(2)} \in (\mathrm{span}(\mathcal{K}))^\perp \subseteq \mathcal{K}^*$ implies that $\breve{X}^{(2)}$ is feasible for ($\mathrm{P}_{\mathrm{conic}}$), and by (3.6b),

  $$b^\top\tilde{y} = v_{\mathrm{P}_{\mathrm{conic}}} \leq v_{\mathrm{D}_{\mathrm{conic}}} \leq b^\top\tilde{y} + \langle \tilde{C}, \breve{X}\rangle_{\mathbb{V}} = b^\top\tilde{y},$$

  where the last equality follows from $\tilde{C} \in \mathcal{K}$.

- *Case 2.* $\text{range}(\mathcal{A}^*) \cap \text{span}(\mathcal{K}) \neq \{0\}$ (i.e., $r := \dim(\text{range}(\mathcal{A}^*) \cap \text{span}(\mathcal{K})) > 0$)

  Let $\mathcal{P} : \mathbb{R}^r \to \mathbb{R}^m$ be a one-one linear map such that

$$\text{range}(\mathcal{A}^*\mathcal{P}) = \text{range}(\mathcal{A}^*) \cap \text{span}(\mathcal{K}). \tag{3.7}$$

  Then using $\tilde{C} \in \mathcal{K}$,

$$\begin{aligned} \tilde{C} - \mathcal{A}^*y \in \mathcal{K} &\iff \tilde{C} - \mathcal{A}^*y \in \mathcal{K},\ \mathcal{A}^*y \in \text{span}(\mathcal{K}) \\ &\iff \tilde{C} - \mathcal{A}^*y \in \mathcal{K},\ y = \mathcal{P}w \ \text{ for some } w \in \mathbb{R}^r. \end{aligned}$$

  Hence following (3.6a),

$$v_{\text{P}_{\text{conic}}} = b^\top \tilde{y} + \sup_w \left\{ (\mathcal{P}^*b)^\top w : \tilde{C} - \mathcal{A}^*\mathcal{P}w \in \mathcal{K} \right\}.$$

  If $\mathcal{P}^*b = 0$, then $v_{\text{P}_{\text{conic}}} = b^\top \tilde{y}$. On the other hand,

$$v_{\text{P}_{\text{conic}}} \leq v_{\text{D}_{\text{conic}}} \leq b^\top \tilde{y} + \inf_X \left\{ \langle \tilde{C}, X \rangle_{\mathbb{V}} : \mathcal{P}^*\mathcal{A}(X) = \mathcal{P}^*b,\ X \in \mathcal{K}^* \right\} \leq b^\top \tilde{y},$$

  where the last inequality follows because $X = 0$ is a feasible solution. Hence $v_{\text{P}_{\text{conic}}} = b^\top \tilde{y}$, and any feasible solution of (D_{conic}) is an optimal solution. Therefore strong duality holds for (P_{conic}).

  It remains to consider the case where $\mathcal{P}^*b \neq 0$. In this case, the set

$$\mathcal{S} := \left\{ Z \in \mathbb{V} : Z = \tilde{C} - \mathcal{A}^*\mathcal{P}w,\ (\mathcal{P}^*b)^\top w > v_{\text{P}_{\text{conic}}} - b^\top \tilde{y} \right\}$$

  is nonempty, and $\mathcal{K} \cap \mathcal{S} = \varnothing$. Observe that $\mathcal{S} \subseteq \text{span}(\mathcal{K})$, as $\text{range}(\mathcal{A}^*\mathcal{P}) \subseteq \text{span}(\mathcal{K})$ and $\tilde{C} \in \mathcal{K}$. By separation theorem (Theorem 2.1.2), there exist $\alpha \in \mathbb{R}$ and a nonzero $\bar{X} \in \text{span}(\mathcal{K})$ such that

$$\langle \bar{X}, Z \rangle_{\mathbb{V}} \geq \alpha \geq \langle \bar{X}, Y \rangle_{\mathbb{V}},\ \forall\, Z \in \mathcal{K},\ Y \in \mathcal{S}. \tag{3.8}$$

  Since $\mathcal{K}$ is a cone, we must have $\langle \bar{X}, Z \rangle_{\mathbb{V}} \geq 0$ for all $Z \in \mathcal{K}$. Hence $\alpha \leq 0$ and $\bar{X} \in \mathcal{K}^*$. We also get from (3.8) and the definition of $\mathcal{S}$ that for any $w \in \mathbb{R}^r$,

$$(\mathcal{P}^*b)^\top w > v_{\text{P}_{\text{conic}}} - b^\top \tilde{y} \implies (\mathcal{P}^*\mathcal{A}(\bar{X}))^\top w \geq \langle \tilde{C}, \bar{X} \rangle_{\mathbb{V}} - \alpha. \tag{3.9}$$

  Then we get that $\mathcal{P}^*\mathcal{A}(\bar{X}) = \beta \mathcal{P}^*b$ for some $\beta \in \mathbb{R}$.[1]

[1] Write $\mathcal{P}^*\mathcal{A}(\bar{X}) = \beta\mathcal{P}^*b + z$, where $z \in \{\mathcal{P}^*b\}^\perp$. Fix any w with $(\mathcal{P}^*b)^\top w > v_{\text{P}_{\text{conic}}} - b^\top \tilde{y}$. For all $\kappa > 0$, we have $(\mathcal{P}^*b)^\top (w - \kappa z) = (\mathcal{P}^*b)^\top w > v_{\text{P}_{\text{conic}}} - b^\top \tilde{y}$, so by (3.9) and $(\mathcal{P}^*\mathcal{A}(\bar{X}))^\top z = \|z\|^2$,

$$\langle \tilde{C}, \bar{X} \rangle_{\mathbb{V}} - \alpha \leq (\mathcal{P}^*\mathcal{A}(\bar{X}))^\top (w - \kappa z) = (\mathcal{P}^*\mathcal{A}(\bar{X}))^\top w - \kappa \|z\|^2.$$

This inequality implies that $z = 0$, for otherwise the right hand side goes to $-\infty$ as $\kappa \to \infty$.

We show that $\beta > 0$. By (3.8), $\beta\kappa\|\mathcal{P}^*b\|^2 \geq \langle \tilde{C}, \bar{X}\rangle_{\mathbb{V}} - \alpha$ for all sufficiently large $\kappa > 0$, so β must be nonnegative. If $\beta = 0$, then $\langle \tilde{C}, \bar{X}\rangle_{\mathbb{V}} \leq \alpha \leq 0$. On the other hand, $\tilde{C} \in \mathrm{ri}(\mathcal{K})$ and $0 \neq \bar{X} \in \mathrm{span}(\mathcal{K}) \cap \mathcal{K}^*$, so $\langle \tilde{C}, \bar{X}\rangle_{\mathbb{V}} > 0$. This contradiction shows that we cannot have $\beta = 0$. Hence $\beta > 0$.

Now we construct an optimal solution of ($\mathrm{D}_{\mathrm{conic}}$). As mentioned in Case 1, $b = \mathcal{A}(\breve{X})$ for some $\breve{X} \in \mathbb{V}$. Then $\mathcal{P}^*\mathcal{A}(\bar{X}) = \beta\mathcal{P}^*b$ implies $\mathcal{P}^*\mathcal{A}(\bar{X} - \beta\breve{X}) = 0$, i.e., $\frac{1}{\beta}\bar{X} - \breve{X} \in \ker(\mathcal{P}^*\mathcal{A}) = \ker(\mathcal{A}) + \mathrm{span}(\mathcal{K})^\perp$, by (3.7). Hence there exist $\breve{X}^{(1)} \in \ker(\mathcal{A})$ and $\breve{X}^{(2)} \in (\mathrm{span}(\mathcal{K}))^\perp$ such that $\frac{1}{\beta}\bar{X} - \breve{X} = \breve{X}^{(1)} + \breve{X}^{(2)}$. Define $\bar{\bar{X}} := \frac{1}{\beta}\bar{X} - \breve{X}^{(2)} = \breve{X}^{(1)} + \breve{X}$. Then $\bar{\bar{X}}$ is feasible for ($\mathrm{D}_{\mathrm{conic}}$), since $\mathcal{A}(\bar{\bar{X}}) = \mathcal{A}(\breve{X}^{(1)} + \breve{X}) = b$ and $\bar{\bar{X}} \in \mathcal{K}^*$. Moreover, $\langle \tilde{C}, \bar{\bar{X}}\rangle_{\mathbb{V}} = \frac{1}{\beta}\langle \tilde{C}, \bar{X}\rangle_{\mathbb{V}}$. Now let $\left\{ Z^{(k)} = \tilde{C} - \mathcal{A}^*\mathcal{P}w^{(k)} \right\}_k$ be a sequence in $\mathcal{S}$ such that $\lim_k (\mathcal{P}^*b)^\top w^{(k)} = v_{\mathrm{P}_{\mathrm{conic}}} - b^\top\tilde{y}$. Then by (3.9), for all k,

$$\langle \tilde{C}, \bar{\bar{X}}\rangle_{\mathbb{V}} = \left\langle \tilde{C}, \frac{1}{\beta}\bar{X} \right\rangle_{\mathbb{V}} - \alpha \leq \frac{1}{\beta}(\mathcal{P}^*\mathcal{A}(\bar{X}))^\top w^{(k)} = (\mathcal{P}^*b)^\top w^{(k)},$$

and taking $k \to \infty$ we have $\langle \tilde{C}, \bar{\bar{X}}\rangle_{\mathbb{V}} \leq v_{\mathrm{P}_{\mathrm{conic}}} - b^\top\tilde{y}$. On the other hand, $v_{\mathrm{P}_{\mathrm{conic}}} \leq v_{\mathrm{D}_{\mathrm{conic}}} \leq b^\top\tilde{y} + \langle \tilde{C}, \bar{\bar{X}}\rangle_{\mathbb{V}}$ by (3.6b) and weak duality. Therefore we have $v_{\mathrm{P}_{\mathrm{conic}}} = v_{\mathrm{D}_{\mathrm{conic}}} = b^\top\tilde{y} + \langle \tilde{C}, \bar{\bar{X}}\rangle_{\mathbb{V}}$, and $\bar{\bar{X}}$ is an optimal solution of (3.6b) and of ($\mathrm{D}_{\mathrm{conic}}$) too.

Therefore in both cases $v_{\mathrm{P}_{\mathrm{conic}}} = v_{\mathrm{D}_{\mathrm{conic}}}$ and $v_{\mathrm{D}_{\mathrm{conic}}}$ is attained. □

Existence of the central path in SDP

For simplicity, we discuss the third application limited to SDP. Other than guaranteeing strong duality, the Slater condition is often assumed to hold for both (P) and (D), so that the central path is well-defined. The *central path* of (P)-(D) is defined as the set $\mathcal{C} := \{(X(\mu), y(\mu), Z(\mu)) : \mu > 0\}$, where $(X(\mu), y(\mu), Z(\mu))$ is the solution of the parametrized system

$$\begin{aligned} Z = C - \mathcal{A}^*y, \quad & Z \succeq 0, \\ \mathcal{A}(X) = b, \quad & X \succeq 0, \\ XZ = \mu I. & \end{aligned} \tag{3.10}$$

Theorem 3.3.4. *(see, e.g., [88, Theorem 5.2]) Suppose that the Slater condition holds for* (P) *and* (D)*. Then for any $\mu > 0$, the system* (3.10) *has a unique solution $(X(\mu), y(\mu), Z(\mu))$, and the central path $\mathcal{C}$ is well-defined.*

Conversely, if for some $\mu > 0$ the system (3.10) has a solution (X, y, Z), then $X = \mu Z^{-1} \succ 0$ is a Slater point for (D), and $Z = \mu X^{-1} \succ 0$ so (y, Z) is a Slater point for (P). Hence the nonemptyness of $\mathcal{C}$ implies that both (P) and (D) satisfy the Slater condition.

3.3.2 Minimal faces of semidefinite programs

A way to understanding the Slater condition is via the notion of minimal faces. The *minimal face of* $(\mathrm{P}_{\mathrm{conic}})$ is the minimal face of $\mathcal{K}$ containing the feasible region of $(\mathrm{P}_{\mathrm{conic}})$, i.e., by Definition 2.2.4:

$$\mathrm{face}\,(\mathrm{P}_{\mathrm{conic}}) := \mathrm{face}(\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K}) = \bigcap \left\{ \mathcal{F} : \mathcal{F} \trianglelefteq \mathcal{K},\ \mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}} \subseteq \mathcal{F} \right\}.$$

By definition, $\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}} \subseteq \mathrm{face}(\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})$. Moreover, from Proposition 2.2.5, $(\mathrm{P}_{\mathrm{conic}})$ satisfies the Slater condition if and only if $\mathrm{face}\,(\mathrm{P}_{\mathrm{conic}}) = \mathcal{K}$. The importance of the minimal face of $(\mathrm{P}_{\mathrm{conic}})$ is that restricting the feasible slack to be contained in $\mathrm{face}\,(\mathrm{P}_{\mathrm{conic}})$ results in an equivalent program (with the same optimal value) for which the Slater condition holds.

Theorem 3.3.5. *[20] Suppose that* $(\mathrm{P}_{\mathrm{conic}})$ *is feasible. Then* $v_{\mathrm{P}_{\mathrm{conic}}} = v^{\mathrm{reg}}_{\mathrm{P}_{\mathrm{conic}}}$*, where*

$$v^{\mathrm{reg}}_{\mathrm{P}_{\mathrm{conic}}} := \sup_y \left\{ b^\top y : C - \mathcal{A}^* y \in \mathrm{face}(\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K}) \right\}. \tag{3.11}$$

Moreover, if the optimal value $v_{\mathrm{P}_{\mathrm{conic}}}$ *is finite, then strong duality holds for* (7.25)*, i.e.,* $v^{\mathrm{reg}}_{\mathrm{P}_{\mathrm{conic}}} = v^{\mathrm{reg}}_{\mathrm{D}_{\mathrm{conic}}}$*, where* $v^{\mathrm{reg}}_{\mathrm{D}_{\mathrm{conic}}}$ *is the optimal value of the dual of* (7.25)*, given by*

$$v^{\mathrm{reg}}_{\mathrm{D}_{\mathrm{conic}}} := \inf_W \left\{ \langle C, X\rangle_{\mathbb{V}} : \mathcal{A}(X) = b,\ X \in \left(\mathrm{face}(\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K})\right)^* \right\}, \tag{3.12}$$

and $v^{\mathrm{reg}}_{\mathrm{D}_{\mathrm{conic}}}$ *is attained.*

Proof. Since $\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}} \subseteq \mathrm{face}(\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K}) \subseteq \mathcal{K}$, for any $y \in \mathbb{R}^m$,

$$C - \mathcal{A}^* y \in \mathcal{K} \iff C - \mathcal{A}^* y \in \mathrm{face}(\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K}).$$

This shows that $v_{\mathrm{P}_{\mathrm{conic}}} = v^{\mathrm{reg}}_{\mathrm{P}_{\mathrm{conic}}}$.

Now assume that $v_{\mathrm{P}_{\mathrm{conic}}}$ is finite. The feasibility of $(\mathrm{P}_{\mathrm{conic}})$ and Proposition 2.2.5 imply that $\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}} \cap \mathrm{ri}(\mathrm{face}(\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}}, \mathcal{K}))$ is nonempty. Therefore, by strong duality theorem (Theorem 3.3.3), $v^{\mathrm{reg}}_{\mathrm{P}_{\mathrm{conic}}} = v^{\mathrm{reg}}_{\mathrm{D}_{\mathrm{conic}}}$ and $v^{\mathrm{reg}}_{\mathrm{D}_{\mathrm{conic}}}$ is attained. □

In the case of SDP, we already see from Proposition 2.2.14 that if (P) does not satisfy the Slater condition, i.e., if $\mathrm{face}(\mathcal{F}^Z_{\mathrm{P}}, \mathbb{S}^n_+) \neq \mathbb{S}^n_+$ (by Proposition 2.2.5), then either $\mathcal{F}^Z_{\mathrm{P}} = \{0\}$ or $\mathrm{face}(\mathcal{F}^Z_{\mathrm{P}}, \mathbb{S}^n_+) = Q\mathbb{S}^r_+ Q^\top$ for some $Q \in \mathbb{R}^{n\times r}$ satisfying $Q^\top Q = I$ and $0 < r < n$. In the latter case, (7.25)-(7.24) would be equivalent to

$$\sup_y \left\{ b^\top y : Q^\top (C - \mathcal{A}^* y) Q \in \mathbb{S}^r_+ \right\}, \tag{3.13}$$

$$\inf_X \left\{ \langle Q^\top C Q, X\rangle : \mathcal{A}(QXQ^\top) = b,\ X \in \mathbb{S}^r_+ \right\}, \tag{3.14}$$

i.e., the feasible slacks lie in a "smaller" positive semidefinite cone $\mathbb{S}^r_+$ with $r < n$. Reducing (P) to the equivalent SDP (3.13) regularizes the SDP (P) and also reduces the number of variables.

3.3.3 Characterizations of the Slater condition

In this section we mention some equivalent conditions of the Slater condition on the general conic program ($\mathrm{P_{conic}}$). We first state the characterizations of the Slater condition for the general conic program ($\mathrm{P_{conic}}$). Define

$$\mathcal{A}_C : \mathbb{V} \to \mathbb{R}^{m+1} : X \mapsto \begin{pmatrix} \mathcal{A}(X) \\ \langle C, X \rangle \end{pmatrix}, \tag{3.15}$$

$$\bar{\mathcal{L}} := \mathrm{range}(\mathcal{A}_C^*) \subseteq \mathbb{V}. \tag{3.16}$$

We will use the same notation $\mathcal{A}_C$ and $\bar{\mathcal{L}}$ for SDP, i.e., when $\mathbb{V} = \mathbb{S}^n$ and $\mathcal{K} = \mathbb{S}^n_+$. If ($\mathrm{P_{conic}}$) is assumed to be feasible, then the Slater condition is equivalent to a nonempty intersection between $\mathrm{ri}(\mathcal{K})$ and the linear subspace $\bar{\mathcal{L}}$.

Lemma 3.3.6. *Suppose that* ($\mathrm{P_{conic}}$) *is feasible. Then*

$$\mathrm{face}\,(\mathrm{P_{conic}}) = \mathrm{face}(\bar{\mathcal{L}} \cap \mathcal{K},\, \mathcal{K}), \tag{3.17}$$

and ($\mathrm{P_{conic}}$) *satisfies the Slater condition if and only if* $\bar{\mathcal{L}} \cap \mathrm{ri}(\mathcal{K}) = \varnothing$.

Proof. We first prove (3.17). Since $\mathcal{F}^Z_{\mathrm{P_{conic}}} \subseteq \bar{\mathcal{L}} \cap \mathcal{K}$, we have $\mathrm{face}\,(\mathrm{P_{conic}}) \subseteq \mathrm{face}(\bar{\mathcal{L}} \cap \mathcal{K},\, \mathcal{K})$.[2] Conversely, for any face $\mathcal{F} \trianglelefteq \mathcal{K}$ containing $\mathcal{F}^Z_{\mathrm{P_{conic}}}$, we show that $\bar{\mathcal{L}} \cap \mathcal{K} \subseteq \mathcal{F}$. Let $Z \in \bar{\mathcal{L}} \cap \mathcal{K}$. Fix $\beta > 0$ and $z \in \mathbb{R}^{m+1}$ such that $\beta Z = \mathcal{A}_C^* z$, $z = \begin{pmatrix} -\tilde{y} \\ \alpha \end{pmatrix}$ and $\alpha \in \{-1, 0, 1\}$. We consider the different possible values of α.

- If $\alpha = 1$, then $\beta Z = C - \mathcal{A}^*\tilde{y} \in \mathcal{F}^Z_{\mathrm{P_{conic}}} \subseteq \mathcal{F}$.
- If $\alpha = 0$, then $\beta Z + (C - \mathcal{A}^*\hat{y}) = C - \mathcal{A}^*(\tilde{y} + \hat{y}) \in \mathcal{F}^Z_{\mathrm{P_{conic}}} \subseteq \mathcal{F}$, where $C - \mathcal{A}^*\hat{y}$ is any element of $\mathcal{F}^Z_{\mathrm{P_{conic}}}$. But $\beta Z, C - \mathcal{A}^*\hat{y} \in \mathcal{K}$, so $\mathcal{F} \trianglelefteq \mathcal{K}$ implies $\beta Z \in \mathcal{F}$.
- If $\alpha = -1$, then $\beta Z + 2(C - \mathcal{A}^*\hat{y}) = C - \mathcal{A}^*(\tilde{y} + 2\hat{y}) \in \mathcal{F}^Z_{\mathrm{P_{conic}}} \subseteq \mathcal{F}$. But $\beta Z, 2(C - \mathcal{A}^*\hat{y}) \in \mathcal{K}$, so $\beta Z \in \mathcal{F}$.

Therefore, in all cases $\beta Z \in \mathcal{F}$, which is a cone, so $Z \in \mathcal{F}$. Then we get $\bar{\mathcal{L}} \cap \mathcal{K} \subseteq \mathcal{F}$. Since $\mathcal{F} \trianglelefteq \mathcal{K}$ containing $\mathcal{F}^Z_{\mathrm{P_{conic}}}$ is arbitrary, we have $\bar{\mathcal{L}} \cap \mathcal{K} \subseteq \mathrm{face}\,(\mathrm{P_{conic}})$. Consequently, $\mathrm{face}(\bar{\mathcal{L}} \cap \mathcal{K}, \mathcal{K}) = \mathrm{face}\,(\mathrm{P_{conic}})$.

[2] For any face $\mathcal{F} \trianglelefteq \mathcal{K}$ such that $\bar{\mathcal{L}} \cap \mathcal{K} \subseteq \mathcal{F}$, we have $\mathcal{F}^Z_{\mathrm{P_{conic}}} \subseteq \mathcal{F}$. So $\mathcal{F}^Z_{\mathrm{P_{conic}}} \subseteq \mathrm{face}(\bar{\mathcal{L}} \cap \mathcal{K}, \mathcal{K})$, implying $\mathrm{face}\,(\mathrm{P_{conic}}) \subseteq \mathrm{face}(\bar{\mathcal{L}} \cap \mathcal{K}, \mathcal{K})$.

For the second claim, it suffices to note that, by Proposition 2.2.5, ($\mathrm{P}_{\mathrm{conic}}$) satisfies the Slater condition (i.e., $\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}} \cap \mathrm{ri}(\mathcal{K}) \neq \varnothing$) if and only if $\mathrm{face}(\bar{\mathcal{L}} \cap \mathcal{K}, \mathcal{K}) = \mathrm{face}\,(\mathrm{P}_{\mathrm{conic}}) = \mathcal{K}$, if and only if $\bar{\mathcal{L}} \cap \mathrm{ri}(\mathcal{K}) = \varnothing$. □

By Lemma 3.3.6, the Slater condition holds for ($\mathrm{P}_{\mathrm{conic}}$) if and only if one of the following equivalent conditions holds:

- The convex set $\mathcal{F}^Z_{\mathrm{P}} \subseteq \mathcal{K}$ has a nonempty intersection with $\mathrm{ri}(\mathcal{K})$.
- The linear subspace $\mathcal{L}$ has a nonempty intersection with $\mathrm{ri}(\mathcal{K})$.
- $\mathcal{F}^Z_{\mathrm{P}}$ is not fully contained within a proper face of $\mathcal{K}$.

We first state an equivalent condition for a nonempty convex subset $\mathcal{S}$ of a closed convex cone $\mathcal{K}$ to have a nonempty intersection with $\mathrm{ri}(\mathcal{K})$.

Theorem 3.3.7. *Let $\mathcal{K}$ be a nonempty closed convex cone in $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$ and $\mathcal{S} \subseteq \mathcal{K}$ be a nonempty convex set. Then*

$$\mathcal{S} \cap \mathrm{ri}(\mathcal{K}) \neq \varnothing \iff \mathcal{S}^\perp \cap \mathcal{K}^* \subseteq -\mathcal{K}^*.$$

Proof. First observe that $\mathcal{S}^\perp \cap \mathcal{K}^* \not\subseteq -\mathcal{K}^*$ if and only if

$$\exists\, d \in \mathcal{S}^\perp \cap \mathcal{K}^*,\, y \in \mathcal{K} \;\text{ s.t. }\; \langle d, y\rangle_{\mathbb{V}} > 0. \tag{3.18}$$

Suppose that $z \in \mathcal{S} \cap \mathrm{ri}(\mathcal{K})$. If, on the contrary, (3.18) holds, then let $\alpha \in (0,1)$ and $\hat{y} \in \mathcal{K}$ satisfy $z = \alpha\hat{y} + (1-\alpha)y$. Then $d \in \mathcal{S}^\perp$ implies that $\alpha\langle d, \hat{y}\rangle_{\mathbb{V}} + (1-\alpha)\langle d, y\rangle_{\mathbb{V}} = \langle d, z\rangle_{\mathbb{V}} = 0$, so $\langle d, \hat{y}\rangle_{\mathbb{V}} = -\frac{1-\alpha}{\alpha}\langle d, y\rangle_{\mathbb{V}} < 0$. This contradicts with the facts that $d \in \mathcal{K}^*$ and $\hat{y} \in \mathcal{K}$. Hence $\mathcal{S}^\perp \cap \mathcal{K}^* \subseteq -\mathcal{K}^*$.

Conversely, suppose that $\mathcal{S} \cap \mathrm{ri}(\mathcal{K}) = \varnothing$. By separation theorem (Theorem 2.1.2), there exist $\beta \in \mathbb{R}$ and $0 \neq d \in \mathbb{V}$ such that

$$\langle d, x\rangle_{\mathbb{V}} \geq \beta \geq \langle d, y\rangle_{\mathbb{V}},\ \forall\, x \in \mathcal{K},\ y \in \mathcal{S}, \quad \text{and} \quad \sup_{x\in\mathcal{K}} \langle d, x\rangle_{\mathbb{V}} > \beta. \tag{3.19}$$

Since $\mathcal{K}$ is a closed cone, we must have $d \in \mathcal{K}^*$ and $\beta \leq 0$. On the other hand, since $\mathcal{S} \subseteq \mathcal{K}$, $0 \leq \langle d, y\rangle_{\mathbb{V}} = \beta$ for all $y \in \mathcal{S}$. Hence $\beta = 0$ and $d \in \mathcal{S}^\perp$. Finally, since $\sup_{x\in\mathcal{K}}\langle d, x\rangle_{\mathbb{V}} > 0$, (3.18) holds, i.e., $\mathcal{S}^\perp \cap \mathcal{K}^* \subseteq \mathcal{K}^*$ does not hold. □

Remark. Any nonzero d satisfying (3.19) defines a proper face of $\mathcal{K}$ containing $\mathcal{S}$: $\mathcal{S} \subseteq \mathcal{K} \cap \{d\}^\perp \lhd \mathcal{K}$.

As a corollary, we apply Theorem 3.3.7 to the case where $\mathcal{K}$ is a face of the cone of positive semidefinite matrices.

Corollary 3.3.8. *Let $Q \in \mathbb{R}^{n\times\bar{n}}$ have orthonormal columns, and $\varnothing \neq \mathcal{S} \subseteq Q\mathbb{S}^{\bar{n}}_{+}Q$. Then*

$$\mathcal{S} \cap Q\mathbb{S}^{n}_{++}Q^{\top} = \varnothing \iff (Q^{\top}\mathcal{S}Q)^{\perp} \cap \mathbb{S}^{\bar{n}}_{+} \neq \{0\}\,.$$

Proof. Let $\bar{\mathcal{S}} := Q^{\top}\mathcal{S}Q$. Then $\mathcal{S} \subseteq Q\mathbb{S}^{\bar{n}}_{+}Q^{\top}$ implies that $\mathcal{S} = QQ^{\top}\mathcal{S}QQ^{\top} = Q\bar{\mathcal{S}}Q^{\top}$.

Using (3.18), $\mathcal{S} \cap Q\mathbb{S}^{n}_{++}Q^{\top} = \varnothing$ if and only if there exists $D \in \mathbb{S}^{n}$ such that

$$D \in \mathcal{S}^{\perp}, \quad 0 \neq Q^{\top}DQ \in \mathbb{S}^{\bar{n}}_{+}. \tag{3.20}$$

Let $D \in \mathbb{S}^{n}$ satisfy (3.20). Then $0 \neq \bar{D} := Q^{\top}DQ \in \mathbb{S}^{\bar{n}}_{+}$. Fix any $\bar{S} \in \bar{\mathcal{S}}$. Then $Q\bar{S}Q^{\top} \in \mathcal{S}$ and $\langle \bar{D}, \bar{S}\rangle = \langle D, Q\bar{S}Q^{\top}\rangle = 0$. Hence $\bar{D} \in \bar{\mathcal{S}}^{\perp} \cap \mathbb{S}^{\bar{n}}_{+}$. Conversely, let $0 \neq \bar{D} \in \bar{\mathcal{S}}^{\perp} \cap \mathbb{S}^{\bar{n}}_{+}$. Let $D := Q\bar{D}Q^{\top}$. Then $Q^{\top}DQ = \bar{D} \neq 0$ and is positive semidefinite. Also, for any $S \in \mathcal{S} = Q\bar{\mathcal{S}}Q^{\top}$, we have $Q^{\top}SQ \in \bar{\mathcal{S}}$ so $\langle D, S\rangle = \langle \bar{D}, Q^{\top}SQ\rangle = 0$. Therefore D satisfies (3.20). □

Now we state an equivalent condition for a linear subspace to have a nonempty intersection with the relative interior of a closed convex cone.

Theorem 3.3.9. *([62, Corollary 2] and [84, Corollary 2.2]) Let $\mathcal{K}$ be a nonempty closed convex cone and $\mathcal{L}$ be a linear subspace in $(\mathbb{V}, \langle\cdot,\cdot\rangle_{\mathbb{V}})$. Then*

$$\mathcal{L} \cap \mathrm{ri}(\mathcal{K}) \neq \varnothing \iff \mathcal{L}^{\perp} \cap \mathcal{K}^{*} \subseteq -\mathcal{K}^{*}.$$

Proof. First observe that $\mathcal{L}^{\perp} \cap \mathcal{K}^{*} \not\subseteq -\mathcal{K}^{*}$ if and only if

$$\exists\, d \in \mathcal{L}^{\perp} \cap \mathcal{K}^{*},\ y \in \mathcal{K} \ \text{ s.t. } \ \langle d, y\rangle_{\mathbb{V}} > 0. \tag{3.21}$$

Suppose that $z \in \mathcal{L} \cap \mathrm{ri}(\mathcal{K})$. If, on the contrary, (3.21) holds, then let $\alpha \in (0,1)$ and $\hat{y} \in \mathcal{K}$ satisfy $z = \alpha\hat{y} + (1-\alpha)y$. Then $d \in \mathcal{L}^{\perp}$ implies that $\alpha\langle d, \hat{y}\rangle_{\mathbb{V}} + (1-\alpha)\langle d, y\rangle_{\mathbb{V}} = \langle d, z\rangle_{\mathbb{V}} = 0$, so $\langle d, \hat{y}\rangle_{\mathbb{V}} < 0$. This contradicts with the facts that $d \in \mathcal{K}^{*}$ and $\hat{y} \in \mathcal{K}$. Hence $\mathcal{L}^{\perp} \cap \mathcal{K}^{*} \subseteq -\mathcal{K}^{*}$.

Conversely, suppose that $\mathcal{L} \cap \mathrm{ri}(\mathcal{K}) = \varnothing$. By separation theorem (Theorem 2.1.1), there exist $\beta \in \mathbb{R}$ and $0 \neq d \in \mathbb{V}$ such that

$$\langle d, x\rangle_{\mathbb{V}} = \beta,\ \forall\, x \in \mathcal{L}, \quad \text{and} \quad \langle d, y\rangle_{\mathbb{V}} > \beta,\ \forall\, y \in \mathrm{ri}(\mathcal{K}).$$

Since $\mathcal{L}$ contains 0, we get $\beta = 0$. This implies that $d \in (\mathcal{L}^{\perp} \cap \mathcal{K}^{*})\backslash(-\mathcal{K}^{*})$. □

As a special case of Theorem 3.3.9, if $\mathcal{K}$ is a proper cone and $\mathcal{L}$ is a linear subspace, the $\mathcal{L} \cap \mathrm{int}(\mathcal{K}) \neq \varnothing$ if and only if $\mathcal{L}^\perp \cap \mathcal{K}^* = \{0\}$. Using Theorem 3.3.9, we can obtain a theorem of the alternative for the Slater condition for the SDP (P).

Theorem 3.3.10. *Assume that* (P) *is feasible. Then exactly one of the following holds.*

(1) (P) *satisfies the Slater condition, i.e., there exists* $\tilde{y} \in \mathbb{R}^m$ *such that* $C - \mathcal{A}^*\tilde{y} \succ 0$.

(2) The system

$$\mathcal{A}_C(D) = 0,\ D \succeq 0 \tag{3.22}$$

has a nonzero solution.

Proof. First recall that by Lemma 3.3.6, (1) is equivalent to $\mathrm{range}(\mathcal{A}_C^*) \cap \mathbb{S}^n_{++} \neq \varnothing$, which is equivalent to $\ker(\mathcal{A}_C) \cap \mathbb{S}^n_+ \subseteq -\mathbb{S}^n_+$ by Theorem 3.3.9. But since $\mathbb{S}^n_+$ is a pointed cone,

$$\ker(\mathcal{A}_C) \cap \mathbb{S}^n_+ \subseteq -\mathbb{S}^n_+ \iff \ker(\mathcal{A}_C) \cap \mathbb{S}^n_+ = \{0\}, \quad \text{i.e., } \text{“}\mathcal{A}_C(D) = 0,\ D \succeq 0 \implies D = 0\text{”}.$$

Hence exactly one of the conditions (1) and (2) holds. □

Remark. Theorem 3.3.10 holds for ($\mathrm{P}_{\mathrm{conic}}$) too, provided that $\mathcal{K}$ is a proper cone. In other words, if $\mathcal{K}$ is a proper cone, then ($\mathrm{P}_{\mathrm{conic}}$) fails the Slater condition if and only if

$$\mathcal{A}_C(D) = 0,\ D \in \mathcal{K}^*$$

has a nonzero solution D. If such D exists, then for any feasible solution y of ($\mathrm{P}_{\mathrm{conic}}$), $\langle D, C - \mathcal{A}^*y\rangle_{\mathbb{V}} = 0$, i.e., $C - \mathcal{A}^*y \in \mathcal{K} \cap \{D\}^\perp \lhd \mathcal{K}$.

Note also that any nonzero solution D of (3.22) is a *direction of constancy* for (D) (i.e., for any feasible solution X of (D), for all $\alpha \geq 0$, $X + \alpha D$ is feasible for (D) and has the same objective value as X.

Finally, we state a theorem of the alternative for the Slater condition on (D):

Theorem 3.3.11. *Assume that* (D) *is feasible. Then exactly one of the following holds.*

(1) (D) *satisfies the Slater condition, i.e., there exists* $\tilde{X} \succ 0$ *such that* $\mathcal{A}(\tilde{X}) = b$.

(2) The system

$$0 \neq \mathcal{A}^*v \succeq 0,\ b^\top v = 0 \tag{3.23}$$

has a solution.

Proof. Let $\hat{X} \in \mathbb{S}^n_+$ be a feasible solution of (D), and $\mathcal{V} : \mathbb{S}^n \to \mathbb{R}^s$ be a linear map such that $\text{range}(\mathcal{V}^*) = \ker(\mathcal{A})$. Then (D) satisfies the Slater condition if and only if there exists $\hat{v} \in \mathbb{R}^s$ such that $\hat{X} + \mathcal{V}^* v \succ 0$. Therefore, by Theorem 3.3.10, (D) satisfies the Slater condition if and only if there does not exist $D \in \mathbb{S}^n$ such that

$$0 \neq D \succeq 0,\ \mathcal{V}(D) = 0,\ \langle \hat{X}, D \rangle = 0. \tag{3.24}$$

Note that (3.23) has a solution if and only if (3.24) does. Indeed, since

$$\ker(\mathcal{V}) = \text{range}(\mathcal{V}^*)^\perp = \ker(\mathcal{A})^\perp = \text{range}(\mathcal{A}^*),$$

if D is a solution of (3.24), then $D = \mathcal{A}^* v$ for some $v \in \mathbb{R}^m$. Also, $b = \mathcal{A}(\hat{X})$. Hence $b^\top v = \langle \hat{X}, \mathcal{A}^* v \rangle = 0$, i.e., v is a solution of (3.23). Conversely, if v solves (3.23), then $D := \mathcal{A}^* v$ solves (3.24). Therefore (D) satisfies the Slater condition if and only if (3.23) has no solution. □

Remark. Similar to Theorem 3.3.10, Theorem 3.3.11 holds for general conic program (D_{conic}) if $\mathcal{K}$ is a proper cone (implying $\mathcal{K}^*$ is also a proper cone, see Proposition 2.1.4). In other words, if $\mathcal{K}$ is a pointed cone, then (D_{conic}) fails the Slater condition if and only if

$$0 \neq \mathcal{A}^* v \in \mathcal{K},\ b^\top v = 0 \tag{3.25}$$

has a solution. Moreover, if v is a solution of (3.25), then for any feasible solution X of (D_{conic}), $\langle X, \mathcal{A}^* v \rangle_{\mathbb{V}} = (\mathcal{A}(X))^\top v = b^\top v = 0$, i.e., $X \in \mathcal{K}^* \cap \{\mathcal{A}^* v\}^\perp \triangleleft \mathcal{K}^*$.

Chapter 4

Facial reduction for linear conic programs

In this chapter we describe the facial reduction algorithm for conic programs.

We first make use of the theorems of the alternative stated in Section 3.3.3 and illustrate one iteration of a facial reduction algorithm for general conic programs ($\mathrm{P}_{\mathrm{conic}}$), in Section 4.1. Then we consider the special cases of SOCP in Section 4.2 and SDP in Section 4.3. We will state the results both in terms of the geometric objects (for theoretical purpose) and in terms of the problem data (for algorithmic purpose).

An important step in the facial reduction algorithm is to find an element in the relative interior of the cone of *directions of constancy*, $\mathrm{ri}(\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}})$. (See Lemma 4.1.1.) We introduce a conic program, which we call the *auxiliary problem*, for determining whether $\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}} = \{0\}$. The auxiliary problem and its dual both satisfy the Slater condition; using an interior point method we can find a point in the relative interior of $\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}}$.

4.1 Dual recession direction and the minimal face

In this section, we make use of the results from Section 3.3.3 to identify minimal faces of second order cone programs and semidefinite programs. We also derive the facial reduction algorithm for finding the minimal face of (P) [20, 27, 70, 86, 95].

As mentioned in the remark of Theorem 3.3.10, any nonzero solution D of (3.22) gives a smaller face of $\mathbb{S}^n_+$ containing the feasible slacks. Lemma 4.1.1 takes care of some trivial cases,

when a solution of (3.22) indicates either that ($\mathrm{P}_{\mathrm{conic}}$) satisfies the Slater condition or that the only feasible slack of ($\mathrm{P}_{\mathrm{conic}}$) is 0.

Lemma 4.1.1. *Assume that the conic program* ($\mathrm{P}_{\mathrm{conic}}$) *is feasible and that the cone* $\mathcal{K}$ *is a proper cone. Define*

$$\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}} := \bar{\mathcal{L}}^{\perp} \cap \mathcal{K}^* = \{X \in \mathbb{V} : \mathcal{A}_C(X) = 0,\ X \in \mathcal{K}^*\}. \tag{4.1}$$

Then for any $D \in \mathrm{ri}(\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}})$,

$$\mathrm{face}\,(\mathrm{P}_{\mathrm{conic}}) \trianglelefteq \mathcal{K} \cap \{D\}^{\perp} = \mathcal{K} \cap (\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}})^{\perp} \trianglelefteq \mathcal{K}, \tag{4.2}$$

and

$$v_{\mathrm{P}_{\mathrm{conic}}} = \sup_y \left\{ b^{\top} y : C - \mathcal{A}^* y \in \mathcal{K} \cap \{D\}^{\perp} \right\} = \sup_y \left\{ b^{\top} y : C - \mathcal{A}^* y \in \mathcal{K} \cap (\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}})^{\perp} \right\}. \tag{4.3}$$

Moreover,

(1) ($\mathrm{P}_{\mathrm{conic}}$) *satisfies the Slater condition if and only if* $\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}} = \{0\}$.

(2) If $\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}} = \{0\}$, *then* $\mathrm{face}\,(\mathrm{P}_{\mathrm{conic}}) = \bar{\mathcal{L}} \cap \mathcal{K} = \{0\}$. *Furthermore,*

$$\mathcal{F}^Z_{\mathrm{P}_{\mathrm{conic}}} = \{0\} \iff \mathcal{R}_{\mathrm{D}_{\mathrm{conic}}} \cap \mathrm{int}(\mathcal{K}^*) = \bar{\mathcal{L}}^{\perp} \cap \mathrm{int}(\mathcal{K}^*) \neq \varnothing. \tag{4.4}$$

In other words, exactly one of the following holds:

(I) $\mathrm{ri}(\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}}) = \{0\}$;

(II) $\mathrm{ri}(\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}}) \subseteq \mathrm{int}(\mathcal{K}^*)$;

(III) $\{0\} \neq \mathrm{ri}(\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}})$ *and* $\mathrm{ri}(\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}}) \cap \mathrm{int}(\mathcal{K}^*) = \varnothing$.

Proof. We first prove (4.2). Since $\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}}$ is a convex subset of $\mathcal{K}^*$, by Proposition 2.2.8 we have $\mathcal{K} \cap (\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}})^{\perp} = \mathcal{K} \cap \{D\}^{\perp} \trianglelefteq \mathcal{K}$. Since $D \in \bar{\mathcal{L}}^{\perp}$, we have $\bar{\mathcal{L}} \cap \mathcal{K} \subseteq \{D\}^{\perp} \cap \mathcal{K}$. Hence by Lemma 3.3.6, $\mathrm{face}\,(\mathrm{P}_{\mathrm{conic}}) = \mathrm{face}(\bar{\mathcal{L}} \cap \mathcal{K}, \mathcal{K}) \trianglelefteq \{D\}^{\perp} \cap \mathcal{K}$.

(4.3) follows immediately from (4.2). Item (1) follows directly from Lemma 3.3.6 and Theorem 3.3.9. (See the proof of Theorem 3.3.10.) We prove Item (2). If $\mathcal{F}_{\mathrm{P}_{\mathrm{conic}}} = \{0\}$, then

$$\{0\} \subseteq \bar{\mathcal{L}} \cap \mathcal{K} \subseteq \mathrm{face}(\bar{\mathcal{L}} \cap \mathcal{K}, \mathcal{K}) = \mathrm{face}\,(\mathrm{P}_{\mathrm{conic}}) = \mathrm{face}(\{0\}, \mathcal{K}) = \{0\}.$$

Hence $\bar{\mathcal{L}} \cap \mathcal{K} = \mathrm{face}\,(\mathrm{P}_{\mathrm{conic}}) = \{0\}$. Moreover, by Theorem 3.3.9, we have $\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}} \cap \mathrm{int}(\mathcal{K}) = \bar{\mathcal{L}}^{\perp} \cap \mathrm{int}(\mathcal{K}) \neq \varnothing$. Conversely, suppose that there exists $D \in \mathrm{int}(\mathcal{K}^*)$ such that $\mathcal{A}_C(D) = 0$. For

all $Z = C - \mathcal{A}^* y \in \mathcal{F}^Z_{\mathrm{P_{conic}}}$, we have $\langle D, Z\rangle_{\mathbb{V}} = (\mathcal{A}_C(D))^\top \left(\begin{smallmatrix} -y \\ 1 \end{smallmatrix}\right) = 0$. But $D \in \mathrm{int}(\mathcal{K}^*)$, so $Z = 0$ by Lemma 3.3.2. Hence $\mathcal{R}_{\mathrm{D_{conic}}} \cap \mathrm{int}(\mathcal{K}^*) \neq \varnothing$ implies that $\mathcal{F}^Z_{\mathrm{P_{conic}}} = \{0\}$. This proves (4.4).

The above shows that (I) and $\mathcal{R}_{\mathrm{D_{conic}}} \cap \mathrm{int}(\mathcal{K}^*) \neq \varnothing$ are mutually exclusive, and $\mathcal{R}_{\mathrm{D_{conic}}} \cap \mathrm{int}(\mathcal{K}^*) \neq \varnothing$ if and only if (II) holds, by Proposition 2.2.5. If neither (I) nor (II) holds, i.e., $\mathrm{ri}(\mathcal{R}_{\mathrm{D_{conic}}}) \neq \{0\}$ and $\mathcal{R}_{\mathrm{D_{conic}}} \cap \mathrm{int}(\mathcal{K}^*) = \varnothing$, then $\{0\} \neq \mathrm{face}(\mathcal{R}_{\mathrm{D_{conic}}}, \mathcal{K}^*) \lhd \mathcal{K}^*$. Then by Theorem 2.2.3 and Proposition 2.2.5 (III) must hold. □

Remark. The three possible cases (I)-(III) in Lemma 4.1.1 can be equivalently posed as a description of the face $\mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp$ that appears in (4.2):

$$\mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp = \begin{cases} \mathcal{K} & \text{in Case (I),} \\ \{0\} & \text{in Case (II),} \\ \text{a proper nonzero face of } \mathcal{K} & \text{in Case (III).} \end{cases}$$

In Case (I), ($\mathrm{P_{conic}}$) satisfies the Slater condition, so by Theorem 3.3.3, strong duality holds for ($\mathrm{P_{conic}}$) if $v_{\mathrm{P_{conic}}}$ is finite.

In Case (II), any feasible solution of ($\mathrm{P_{conic}}$) is an optimal solution, and $v_{\mathrm{P_{conic}}} = b^\top (\mathcal{A}^*)^\dagger(C)$. In fact, the dual is always solvable too, under Assumption 3.1, i.e., if $\mathcal{A}(\hat{X}) = b$ for some $\hat{X} \in \mathcal{K}$. Indeed, let $D \in \mathrm{ri}(\mathcal{R}_{\mathrm{D_{conic}}})$. Then $D \in \mathrm{int}(\mathcal{K}^*)$ so $\hat{X} + \gamma D \in \mathcal{K}^*$ for sufficiently large $\gamma \in \mathbb{R}$, and $\langle C, \hat{X} + \gamma D\rangle_{\mathbb{V}} = \langle C, \hat{X}\rangle_{\mathbb{V}}$. This means that ($\mathrm{D_{conic}}$) is feasible and

$$b^\top \Big((\mathcal{A}^*)^\dagger(C)\Big) = v_{\mathrm{P_{conic}}} \leq v_{\mathrm{D_{conic}}} \leq \langle C, \mathcal{A}^\dagger b\rangle_{\mathbb{V}} = \Big((\mathcal{A}^*)^\dagger(C)\Big)^\top b = v_{\mathrm{P_{conic}}},$$

i.e., $\mathcal{A}^\dagger b + \gamma D$ is an optimal solution of ($\mathrm{D_{conic}}$). Note however that the set of optimal solutions of ($\mathrm{D_{conic}}$) is unbounded, echoing the result of Proposition 3.3.1.

The smaller face of $\mathcal{K}$ containing the feasible region, found in Lemma 4.1.1, can be used to formulate ($\mathrm{P_{conic}}$) as a "smaller" equivalent program. We introduce one more technical result that facilitates this, for Case (III) in Lemma 4.1.1.

Lemma 4.1.2. *Assume that the conic program* ($\mathrm{P_{conic}}$) *is feasible and that* $\mathcal{K}$ *is a proper cone. Suppose that* $\mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp$ *does not equal* $\mathcal{K}$ *or* $\{0\}$*, and that* $C \in \mathcal{K}$ *(i.e.,* $y = 0$ *is feasible for* ($\mathrm{P_{conic}}$)*). Then the followings hold.*

(1) If $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp) = \{0\}$*, then* $y = 0$ *is the only feasible solution of* ($\mathrm{P_{conic}}$)*, i.e.,* $\mathcal{F}^Z_{\mathrm{P_{conic}}} = \{C\}$.

(2) If $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp) = \mathrm{range}(\mathcal{A}^*\mathcal{P})$ *for some one-one linear map* $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$ *(where* $\bar{m} > 0$*), then*

$$Z = C - \mathcal{A}^*y \in \mathcal{K},\ y \in \mathbb{R}^m \iff Z = C - \mathcal{A}^*\mathcal{P}v \in \mathcal{K},\ v \in \mathbb{R}^{\bar{m}}. \tag{4.5}$$

In particular, ($\mathrm{P_{conic}}$) *is equivalent to*

$$v_{\mathrm{P_{conic}}} = \sup_v \left\{ (\mathcal{P}^*b)^\top v : C - \mathcal{A}^*\mathcal{P}v \in \mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp \right\}. \tag{4.6}$$

Proof. By (4.2), for any $y \in \mathbb{R}^m$ and $Z = C - \mathcal{A}^*y$,

$$Z \in \mathcal{K} \iff Z \in \mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp \iff Z \in \mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp,\ \mathcal{A}^*y \in \mathrm{span}\left(\mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp\right). \tag{4.7}$$

If $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp) = \{0\}$, then (4.7) implies that $y = 0$ is the only feasible solution of ($\mathrm{P_{conic}}$).Otherwise, $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp) = \mathrm{range}(\mathcal{A}^*\mathcal{P})$ for some one-one linear map $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$, and $y \in \mathcal{F}^y_{\mathrm{P_{conic}}}$ implies $\mathcal{A}^*y \in \mathrm{range}(\mathcal{A}^*\mathcal{P})$, so $y = \mathcal{P}v$ for some $v \in \mathbb{R}^{\bar{m}}$ (since $\mathcal{A}^*$ is one-one). Therefore $Z = C - \mathcal{A}^*\mathcal{P}v \in \mathcal{K}$. This proves (4.5). Finally, using (4.7) and the definition of $\mathcal{P}$,

$$\begin{aligned} v_{\mathrm{P_{conic}}} &= \left\{ b^\top y : Z = C - \mathcal{A}^*y \in \mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp,\ y = \mathcal{P}v \right\} \\ &= \left\{ b^\top(\mathcal{P}v) : Z = C - \mathcal{A}^*\mathcal{P}v \in \mathcal{K} \cap (\mathcal{R}_{\mathrm{D_{conic}}})^\perp \right\}. \end{aligned}$$

This proves (4.6). □

Lemma 4.1.1 suggests that the interesting case is Item (III), in which case for any $D \in \mathrm{ri}(\mathcal{R}_{\mathrm{D_{conic}}})$, the face $\{0\} \neq \mathcal{K} \cap \{D\}^\perp \neq \mathcal{K}$ is a proper face containing the feasible slacks $\mathcal{F}^Z_{\mathrm{P_{conic}}}$. For SDP, we can then use D to write down a "smaller" equivalent program. This idea forms the basis of facial reduction for SDP. We illustrate this idea of reducing conic programs in the case of SOCP in Section 4.2 and SDP in Section 4.3. Finally we define the auxiliary problem, a mixed conic program, for determining whether $\mathcal{R}_{\mathrm{D_{conic}}} = \{0\}$, and if not, finding a nonzero element in $\mathcal{R}_{\mathrm{D_{conic}}}$.

4.2 Single second order cone programs

We consider *single second order cone programs*, i.e., ($\mathrm{P_{conic}}$) with $\mathbb{V} = \mathbb{R}^n$ and $\mathcal{K} = \mathcal{Q}^n$. By Theorem 2.2.13, all proper faces of $\mathcal{Q}^n$ are of the form $\{\alpha z : \alpha \geq 0\}$ for some $z \in \mathbb{R}^n$ with $z_1 = \|z_{2:n}\|$. Given the simple form of the faces, we expect that if a single second order cone program fails the Slater condition, it is relatively easy to identify the minimal face. In fact, we can even solve the optimization problem explicitly in this case.

Theorem 4.2.1. *Assume that the optimization problem*

$$\nu := \sup_y \left\{ b^\top y : z = c - A^\top y \in \mathcal{Q}^n \right\} \tag{4.8}$$

is feasible, where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$ and $c \in \mathbb{R}^n$. Define

$$\mathcal{R} := \left\{ g \in \mathcal{Q}^n : Ag = 0,\ c^\top g = 0 \right\}.$$

Let $d \in \mathrm{ri}(\mathcal{R})$. Then

(1) $d = 0$ if and only if the Slater condition holds for (4.8).

(2) $d_1 > \|d_{2:n}\|$ if and only if $c - A^\top y = 0$ for all feasible y.

(3) If $d_1 = \|d_{2:n}\| > 0$, then

$$z = c - A^\top y \in \mathcal{Q}^n \implies z = \alpha \hat{d}, \quad \text{where } \hat{d} = \begin{pmatrix} d_1 \\ -d_{2:n} \end{pmatrix},\ \alpha = \frac{z_1}{d_1}. \tag{4.9}$$

Suppose that $c = -\frac{c_1}{d_1}\hat{d}$ (i.e., $y = 0$ is a feasible solution of (4.8)*). If $\hat{d} \notin \mathrm{range}(A^\top)$, then $y = 0$ is the only feasible solution of* (4.8). *If $\hat{d} \in \mathrm{range}(A^\top)$, then*

$$\nu = \sup_\gamma \left\{ (b^\top \hat{y})\gamma : \gamma \le \frac{c_1}{d_1} \right\}, \tag{4.10}$$

where $\hat{y}$ satisfies $A^\top \hat{y} = \hat{d}$. Moreover, γ^* solves (4.10) *if and only if $\gamma^* \hat{y}$ solves* (4.8).

Proof. Items (1) and (2) follows from Lemma 4.1.1 (and the fact that $d \in \mathrm{int}(\mathcal{Q}^n)$ if and only if $d_1 > \|d_{2:n}\|$).

It remains to prove Item (3). The condition $d_1 = \|d_{2:n}\| > 0$ implies that (4.8) fails the Slater condition and (by Lemma 4.1.1) $\mathrm{ri}(\mathcal{R}) \cap \mathrm{int}(\mathcal{Q}^n) = \varnothing$. Hence by Lemma 4.1.1, the minimal face of $\mathcal{Q}^n$ containing the feasible region of (4.8) does not equal $\mathcal{Q}^n$ or $\{0\}$. Then by Theorem 2.2.13, the minimal face of $\mathcal{Q}^n$ containing the feasible region of (4.8) equals $\{\alpha z : \alpha \ge 0\}$ for some $z \in \mathbb{R}^n$ with $z_1 = \|z_{2:n}\|$. In fact, for any $z = c - A^\top y \in \mathcal{Q}^n$,

$$0 = d^\top z = d_1 z_1 + (d_{2:n})^\top z_{2:n} \ge d_1 z_1 - \|d_{2:n}\|\|z_{2:n}\| = d_1(z_1 - \|z_{2:n}\|) \ge 0,$$

implying that $z_1 = \|z_{2:n}\|$ and $(d_{2:n})^\top z_{2:n} = -\|d_{2:n}\|\|z_{2:n}\|$. Hence $z_{2:n} = \alpha d_{2:n}$ for some $\alpha \le 0$, and $z_1 = \|z_{2:n}\| = -\alpha\|d_{2:n}\| = -\alpha d_1$. Therefore (4.9) holds.

Now suppose that $c = \frac{c_1}{d_1}\hat{d}$. Then by (4.9),

$$\begin{aligned} c - A^\top y \in \mathcal{Q}^n \quad &\iff \quad c - A^\top y = \beta\hat{d}, \quad \beta \ge 0 \\ &\iff \quad A^\top y = \gamma\hat{d}, \quad \gamma = \tfrac{c_1}{d_1} - \beta \le \tfrac{c_1}{d_1}. \end{aligned} \tag{4.11}$$

If $\hat{d} \notin \text{range}(A^\top)$, then $A^\top y = \gamma \hat{d}$ if and only if $\gamma = 0$, if and only if $y = 0$. Hence the feasible region of (4.8) is the singleton $\{0\}$.

If $\hat{d} = A^\top \hat{y}$ for some (unique) $\hat{y}$, then by (4.11),

$$c - A^\top y \in \mathcal{Q}^n \iff y = \gamma \hat{y},\ \gamma \leq \frac{c_1}{d_1}.$$

Hence (4.10) holds, and γ^* solves (4.10) if and only if $y^* = \gamma^* \hat{y}$ solves (4.8). □

Noting that the single-variable optimization problem (4.10) trivially satisfies the Slater condition, Theorem 4.2.1 shows that any single second order cone program of the form (4.8) requires at most one facial reduction iteration to identify its minimal face.

4.3 Semidefinite programs

Now we consider the semidefinite program (P). Any nonzero solution D of (3.22) is useful not only because it serves as a certificate of the failure of the Slater condition, but also because it can give us a proper face of $\mathbb{S}^n_+$ containing the feasible region of (P).

Theorem 4.3.1. *Assume that* (P) *is feasible. Let* $D \in \text{ri}(\mathcal{R}_\text{D})$*, where*

$$\mathcal{R}_\text{D} = \bar{\mathcal{L}}^\perp \cap \mathbb{S}^n_+ = \{G \in \mathbb{S}^n : \mathcal{A}_C(G) = 0,\ G \succeq 0\}.$$

Then $\text{face}\,(\text{P}) \trianglelefteq \mathbb{S}^n_+ \cap \{D\}^\perp \trianglelefteq \mathbb{S}^n_+$*. Moreover,*

(1) $D = 0$ *if and only if* $\text{face}\,(\text{P}) = \mathbb{S}^n_+$*, i.e., the Slater condition holds for* (P)*.*

(2) $D \succ 0$ *if and only if* $\mathcal{F}^Z_\text{P} = \{0\}$*. Indeed, if* $\mathcal{F}^Z_\text{P} = \{0\}$*, then* $\text{face}\,(\text{P}) = \bar{\mathcal{L}} \cap \mathbb{S}^n_+ = \{0\}$ *too.*

(3) If $D = PD_+P^\top$*, where* $D_+ \in \mathbb{S}^{n-\bar{n}}_{++}$*,* $0 < \bar{n} < n$ *and* $U = \begin{bmatrix} P & Q \end{bmatrix} \in \mathbb{R}^{n\times n}$ *is orthogonal, then*

$$Q^\top \bar{\mathcal{L}} Q \cap \mathbb{S}^n_+ = \begin{bmatrix} 0 & 0 \\ 0 & \bar{\bar{\mathcal{L}}} \cap \mathbb{S}^{\bar{n}}_+ \end{bmatrix}, \quad \textit{and}$$

$$\text{face}\,(\text{P}) = Q \begin{bmatrix} 0 & 0 \\ 0 & \text{face}(\bar{\bar{\mathcal{L}}} \cap \mathbb{S}^{\bar{n}}_+,\ \mathbb{S}^{\bar{n}}_+) \end{bmatrix} Q^\top \trianglelefteq Q\mathbb{S}^{\bar{n}}_+ Q^\top \triangleleft \mathbb{S}^n_+,$$

where $\bar{\bar{\mathcal{L}}} \neq \{0\}$ *is the linear subspace of* $\mathbb{S}^{\bar{n}}$ *determined by* $Q^\top \bar{\mathcal{L}} Q \cap \left[\begin{smallmatrix} 0 & 0 \\ 0 & \mathbb{S}^{\bar{n}} \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & 0 \\ 0 & \bar{\bar{\mathcal{L}}} \end{smallmatrix}\right]$.

Proof. If $\mathcal{A}_C^* z \succeq 0$, then $\langle D, \mathcal{A}_C^* z\rangle = (\mathcal{A}_C(D))^\top z = 0$. Hence $\bar{\mathcal{L}} \cap \mathbb{S}_+^n \subseteq \mathbb{S}_+^n \cap \{D\}^\perp$, and by (3.17) we have face (P) $\subseteq \mathbb{S}_+^n \cap \{D\}^\perp$, which is a face of $\mathbb{S}_+^n$ by Corollary 2.2.15. Since face (P) $\trianglelefteq \mathbb{S}_+^n$, by Proposition 2.2.2 we have face (P) $\trianglelefteq \mathbb{S}_+^n \cap \{D\}^\perp$. Now we consider the three different cases.

(1) $D = 0$ if and only if the system (3.22) has only a zero solution, if and only if (P) satisfies the Slater condition by Theorem 3.3.10, which is equivalent to face (P) $= \mathbb{S}_+^n$ by Proposition 2.2.5.

(2) If $D \succ 0$, then $\varnothing \neq \mathcal{F}_{\mathrm{P}}^Z \subseteq \mathbb{S}_+^n \cap \{D\}^\perp = \{0\}$. Hence $\mathcal{F}_{\mathrm{P}}^Z = \{0\}$. Conversely, if $\mathcal{F}_{\mathrm{P}}^Z = \{0\}$, then face$(\bar{\mathcal{L}} \cap \mathbb{S}_+^n, \mathbb{S}_+^n)$ = face (P) = face$(\{0\}, \mathbb{S}_+^n) = \{0\}$, so $\bar{\mathcal{L}} \cap \mathbb{S}_+^n = \{0\}$. By Theorem 3.3.9, we get $\bar{\mathcal{L}}^\perp \cap \mathbb{S}_{++}^n \neq \varnothing$, i.e., the maximum rank of matrices in $\bar{\mathcal{L}}^\perp \cap \mathbb{S}_+^n$ is n. By Corollary 2.2.18, any matrix in $\mathrm{ri}(\bar{\mathcal{L}}^\perp \cap \mathbb{S}_+^n)$ is of rank n. Hence $D \succ 0$.

(3) Suppose that $D = PD_+P$, where $D_+ \in \mathbb{S}_{++}^{n-\bar{n}}$, $0 < \bar{n} < n$ and $U = \begin{bmatrix} P & Q \end{bmatrix} \in \mathbb{R}^{n\times n}$ is orthogonal. Hence face (P) $\trianglelefteq \mathbb{S}_+^n \cap \{D\}^\perp = Q\mathbb{S}_+^{\bar{n}} Q^\top \triangleleft \mathbb{S}_+^n$. Rotating $\bar{\mathcal{L}} \cap \mathbb{S}_+^n$ using $U^\top \cdot U$:

$$U^\top \bar{\mathcal{L}} U \cap \mathbb{S}_+^n \;=\; U^\top(\bar{\mathcal{L}} \cap \mathbb{S}_+^n)U \;\subseteq\; U^\top(Q\mathbb{S}_+^{\bar{n}}Q^\top)U \;=\; \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{\bar{n}} \end{bmatrix}.$$

This implies that

$$Q^\top \bar{\mathcal{L}} Q \cap \mathbb{S}_+^n \;=\; Q^\top \bar{\mathcal{L}} Q \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{\bar{n}} \end{bmatrix} \;=\; \left(Q^\top \bar{\mathcal{L}} Q \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}^{\bar{n}} \end{bmatrix} \right) \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{\bar{n}} \end{bmatrix} \;=\; \begin{bmatrix} 0 & 0 \\ 0 & \bar{\bar{\mathcal{L}}} \cap \mathbb{S}_+^{\bar{n}} \end{bmatrix}.$$

But both $Q^\top \bar{\mathcal{L}} Q$ and $\left[\begin{smallmatrix} 0 & 0 \\ 0 & \mathbb{S}^{\bar{n}} \end{smallmatrix}\right]$ are linear subspaces of $\mathbb{S}^n$, so the set $\bar{\bar{\mathcal{L}}}$ determined by the subspace intersection $Q^\top \bar{\mathcal{L}} Q \cap \left[\begin{smallmatrix} 0 & 0 \\ 0 & \mathbb{S}^{\bar{n}} \end{smallmatrix}\right] = \left[\begin{smallmatrix} 0 & 0 \\ 0 & \bar{\bar{\mathcal{L}}} \end{smallmatrix}\right]$ is a linear subspace of $\mathbb{S}^{\bar{n}}$. If $\bar{\bar{\mathcal{L}}} = \{0\} \subset \mathbb{S}^{\bar{n}}$, then $Q^\top \bar{\mathcal{L}} Q \cap \mathbb{S}_+^n = \{0\}$, implying that $\bar{\mathcal{L}} \cap \mathbb{S}_+^n = \{0\}$. Then by Item (2) $D \in \mathrm{ri}(\bar{\mathcal{L}}^\perp \cap \mathbb{S}_+^n)$ should have full rank, which is contradictory. Hence $\bar{\bar{\mathcal{L}}}$ cannot be the zero subspace.

It remains to prove the equality in the second claim. Indeed, using Proposition 2.2.17,

$$\begin{aligned}
\operatorname{face}(\mathrm{P}) =& \operatorname{face}(\bar{\mathcal{L}} \cap \mathbb{S}^n_+, \mathbb{S}^n_+) \\
=& QQ^\top \operatorname{face}(\bar{\mathcal{L}} \cap \mathbb{S}^n_+, \mathbb{S}^n_+) QQ^\top \\
=& Q \operatorname{face}(Q^\top \bar{\mathcal{L}} Q \cap \mathbb{S}^n_+, \mathbb{S}^n_+) Q^\top \\
=& Q \operatorname{face}\left(\begin{bmatrix} 0 & 0 \\ 0 & \bar{\bar{\mathcal{L}}} \cap \mathbb{S}^{\bar{n}}_+ \end{bmatrix}, \mathbb{S}^n_+ \right) Q^\top \\
=& Q \operatorname{face}\left(\begin{bmatrix} 0 \\ I_{\bar{n}} \end{bmatrix} \left(\bar{\bar{\mathcal{L}}} \cap \mathbb{S}^{\bar{n}}_+ \right) \begin{bmatrix} 0 & I_{\bar{n}} \end{bmatrix}, \mathbb{S}^n_+ \right) Q^\top \\
=& Q \begin{bmatrix} 0 & 0 \\ 0 & \operatorname{face}(\bar{\bar{\mathcal{L}}} \cap \mathbb{S}^{\bar{n}}_+, \mathbb{S}^{\bar{n}}_+) \end{bmatrix} Q^\top.
\end{aligned}$$

□

Remark. In the third case in Theorem 4.3.1, finding the minimal face of (P) is equivalent to finding the minimal face containing $\bar{\bar{\mathcal{L}}} \cap \mathbb{S}^{\bar{n}}_+$ in the smaller cone $\mathbb{S}^{\bar{n}}_+$, by rotating the cone $\mathbb{S}^n_+$ with $Q^\top \cdot Q$. While the intersection $\operatorname{range}(Q \cdot Q^\top) \cap \operatorname{range}(\mathcal{A}^*_C)$ (or equivalently the linear subspace $\bar{\bar{\mathcal{L}}}$) is guaranteed to be nonzero, the intersection $\operatorname{range}(Q \cdot Q^\top) \cap \operatorname{range}(\mathcal{A}^*)$ could be zero. Here is an example:

$$\sup_y \left\{ y : \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} - y \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \succeq 0 \right\}.$$

It is immediate that $\mathcal{F}^y_{\mathrm{P}} = \{0\}$ and $\mathcal{F}^Z_{\mathrm{P}} = \{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)\}$. Taking $D = \left(\begin{smallmatrix} 0 & 0 \\ 0 & 1 \end{smallmatrix}\right) \in \operatorname{ri}(\mathcal{R}_{\mathrm{D}})$ and $Q = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, we have

$$\operatorname{range}(Q \cdot Q^\top) = \operatorname{span}\left(\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right\} \right), \quad \operatorname{range}(\mathcal{A}^*) = \operatorname{span}\left(\left\{ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \right\} \right),$$

so $\operatorname{range}(Q \cdot Q^\top) \cap \operatorname{range}(\mathcal{A}^*) = \{0\}$ despite $\operatorname{range}(Q \cdot Q^\top) \cap \operatorname{range}(\mathcal{A}^*_C) = \operatorname{span}\left(\{\left(\begin{smallmatrix} 1 & 0 \\ 0 & 0 \end{smallmatrix}\right)\}\right) \neq \{0\}$.

Using Theorem 4.3.1, we can derive a facial reduction algorithm for identifying the minimal face of $\mathbb{S}^n_+$ containing the intersection $\bar{\mathcal{L}} \cap \mathbb{S}^n_+$, outlined in Algorithm 4.1 on Page 55.

Below we rephrase the geometric result from Theorem 4.3.1: Theorem 4.3.1 can be used not only to identify $\operatorname{face}(\mathrm{P}) = \operatorname{face}(\bar{\mathcal{L}} \cap \mathbb{S}^n_+, \mathbb{S}^n_+)$, but also to formulate an equivalent problem to (P) over a "smaller" PSD cone.

Theorem 4.3.2. *Assume that* (P) *is feasible. Let* $D = PD_+P^\top \in \mathcal{R}_{\mathrm{D}}$*, where* $D_+ \in \mathbb{S}^{n-\bar{n}}_{++}$*,* $0 < \bar{n} < n$ *and* $U = \begin{bmatrix} P & Q \end{bmatrix} \in \mathbb{R}^{n \times n}$ *is orthogonal.*

Algorithm 4.1: Identifying face$(\bar{\mathcal{L}} \cap \mathbb{S}^n_+, \mathbb{S}^n_+)$ for any linear subspace $\{0\} \neq \bar{\mathcal{L}} \subseteq \mathbb{S}^n$

Input(*linear subspace* $\{0\} \neq \bar{\mathcal{L}}$ *of* $\mathbb{S}^n$);

find a $D^{(0)} \in \mathrm{ri}(\bar{\mathcal{L}}^\perp \cap \mathbb{S}^n_+)$;

if $D^{(0)} = 0$ **then**

  $d \leftarrow 0$; STOP; % $\bar{\mathcal{L}} \cap \mathbb{S}^n_{++} \neq \varnothing$

endif

if $D^{(0)} \succ 0$ **then**

  $d \leftarrow 0$; STOP; % $\bar{\mathcal{L}} \cap \mathbb{S}^n_+ = \{0\}$

endif

find an orthogonal matrix $Q^{(0)} = \overset{\overset{n-n_1}{\downarrow}}{{}_{n}} \left[\begin{array}{cc} Q_1^{(0)} & Q_2^{(0)} \end{array} \right] \overset{\overset{n_1}{\downarrow}}{} \in \mathbb{R}^{n \times n}$ (with $0 < n_1 < n$) such that

$$D^{(0)} = Q_1^{(0)} D_+^{(0)} (Q_1^{(0)})^\top, \quad D_+^{(0)} \in \mathbb{S}^{n-n_1}_{++};$$

find the linear subspace $\bar{\mathcal{L}}^{(1)} \subseteq \mathbb{S}^{n_1}$ satisfying $(Q^{(0)})^\top \bar{\mathcal{L}} Q^{(0)} \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}^{n_1} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \bar{\mathcal{L}}^{(1)} \end{bmatrix}$;

% $\bar{\mathcal{L}}^{(1)}$ cannot equal to the zero subspace;

for $k = 1, \ldots$ **do**

  find a $D^{(k)} \in \mathrm{ri}\big((\bar{\mathcal{L}}^{(k)})^\perp \cap \mathbb{S}^{n_k}_+\big)$; % $D^{(k)}$ cannot be positive definite

  if $D^{(k)} = 0$ **then**

    $d \leftarrow k$; STOP;

  else

    find an orthogonal matrix $Q^{(k)} = \overset{\overset{n_k-n_{k+1}}{\downarrow}}{{}_{n_k}} \left[\begin{array}{cc} Q_1^{(k)} & Q_2^{(k)} \end{array} \right] \overset{\overset{n_{k+1}}{\downarrow}}{} \in \mathbb{R}^{n_k \times n_k}$ (with $0 < n_{k+1} < n_k$) such that

$$D^{(k)} = Q_1^{(k)} D_+^{(k)} (Q_1^{(k)})^\top, \quad D_+^{(k)} \in \mathbb{S}^{n_k - n_{k+1}}_{++};$$

    find the linear subspace $\bar{\mathcal{L}}^{(k+1)} \subseteq \mathbb{S}^{n_{k+1}}$ satisfying $(Q^{(k)})^\top \bar{\mathcal{L}}^{(k)} Q^{(k)} \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}^{n_{k+1}} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & \bar{\mathcal{L}}^{(k+1)} \end{bmatrix}$;

    % $\bar{\mathcal{L}}^{(k+1)}$ cannot equal to the zero subspace;

  endif

endfor

(1) If $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top) = \{0\}$*, then* $\mathcal{F}_{\mathrm{P}}^y = \{\hat{y}\}$*, where* $\hat{y}$ *is the unique solution to the linear equation*

$$C - QQ^\top CQQ^\top = \mathcal{A}^*\hat{y} - QQ^\top(\mathcal{A}^*\hat{y})QQ^\top. \tag{4.12}$$

(2) If $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top) = \mathrm{range}(\mathcal{A}^*\mathcal{P})$*, where* $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$ *is an one-one linear map, then*

$$v_{\mathrm{P}} = b^\top \bar{y} + \sup_v \left\{ \bar{b}^\top v : \bar{C} - \bar{\mathcal{A}}^* v \succeq 0 \right\}, \tag{4.13}$$

where $\bar{y} \in \mathbb{R}^m$ *is the unique solution of the linear equations*

$$\mathcal{A}^* y - QQ^\top(\mathcal{A}^* y)QQ^\top = C - QQ^\top CQQ^\top, \tag{4.14a}$$

$$\mathcal{P}^* y = 0, \tag{4.14b}$$

and

$$\bar{\mathcal{A}}^*(\cdot) = Q^\top\left(\mathcal{A}^*\mathcal{P}(\cdot)\right)Q, \quad \bar{b} = \mathcal{P}^* b, \quad \bar{C} = Q^\top(C - \mathcal{A}^*\bar{y})Q. \tag{4.15}$$

Moreover, $\bar{\mathcal{A}} : \mathbb{S}^{\bar{n}} \to \mathbb{R}^{\bar{m}}$ *is an onto linear map, and* y *is feasible for* (P) *if and only if* $y - \bar{y} = \mathcal{P}v$ *and* v *is feasible for* (4.13)*.*

Proof. Fix any $\hat{y} \in \mathbb{R}^m$ (e.g., any $\hat{y} \in \mathcal{F}_{\mathrm{P}}^y$) that satisfies $\hat{Z} := C - \mathcal{A}^*\hat{y} \in \mathrm{range}(Q \cdot Q^\top)$. Then

$$v_{\mathrm{P}} = \sup_y \left\{ b^\top y : \hat{Z} - \mathcal{A}^*(y - \hat{y}) \succeq 0 \right\} = b^\top\hat{y} + \sup_y \left\{ b^\top y : \hat{Z} - \mathcal{A}^* y \succeq 0 \right\},$$

and $\hat{Z} - \mathcal{A}^* y \succeq 0$ if and only if $y \in \hat{y} + \mathcal{F}_{\mathrm{P}}^y$. Since $\mathcal{F}_{\mathrm{P}}^Z \subseteq \mathbb{S}_+^n \cap \{D\}^\perp = Q\mathbb{S}_+^{\bar{n}}Q^\top$ and $\hat{Z} = QQ^\top\hat{Z}QQ^\top$,

$$\hat{Z} - \mathcal{A}^* y \in \mathbb{S}_+^n \iff \hat{Z} - \mathcal{A}^* y \in Q\mathbb{S}_+^{\bar{n}}Q \ \text{ and } \ \mathcal{A}^* y \in \mathrm{range}(Q \cdot Q^\top). \tag{4.16}$$

Suppose that $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top) = \{0\}$. From (4.16) we have $\hat{Z} - \mathcal{A}^* y \succeq 0$ if and only if $y = 0$, so $\mathcal{F}_{\mathrm{P}}^y = \{\hat{y}\}$ and $\mathcal{F}_{\mathrm{P}}^Z = \{\hat{Z}\}$. That $\hat{y}$ satisfies (4.12) follows from $\hat{Z} = QQ^\top\hat{Z}QQ^\top$. To see that (4.12) has a unique solution, it suffices to note that the linear map $y \mapsto \mathcal{A}^* y - QQ^\top(\mathcal{A}^* y)QQ^\top$ is one-one: for any $y, \tilde{y} \in \mathbb{R}^m$,

$$\begin{aligned}
&\mathcal{A}^* y - QQ^\top(\mathcal{A}^* y)QQ^\top = \mathcal{A}^*\tilde{y} - QQ^\top(\mathcal{A}^*\tilde{y})QQ^\top \\
\implies &\mathcal{A}^*(y - \tilde{y}) = QQ^\top(\mathcal{A}^*(y - \tilde{y}))QQ^\top \in \mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top) \\
\implies &\mathcal{A}^*(y - \tilde{y}) = 0 \implies y = \tilde{y}.
\end{aligned}$$

Suppose that $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top) = \mathrm{range}(\mathcal{A}^*\mathcal{P})$, where $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$ is an one-one linear map. We first show that (4.14) has a unique solution.

- *Existence of solutions of* (4.14).

  Define the linear map $\mathcal{G}(y) := \mathcal{A}^*y - QQ^\top(\mathcal{A}^*y)QQ^\top$. Then

  - (4.14) reads $\mathcal{G}(y) = C - QQ^\top CQQ^\top$, and
  - $\ker(\mathcal{G}) = \mathrm{range}(\mathcal{P})$.[1]

  Any $y \in \mathcal{F}_{\mathrm{P}}^y$ satisfies (4.14a). Write $y = y_1 + y_2$, where $y_1 \in \mathrm{range}(\mathcal{G}^*)$ and $y_2 \in \ker(\mathcal{G})$. Then $\mathcal{G}(y_1) = \mathcal{G}(y) = C - QQ^\top CQQ^\top$, i.e., y_1 satisfies (4.14). Moreover, y_1 satisfies (4.14b) because $y_1 \in \mathrm{range}(\mathcal{G}^*) = \ker(\mathcal{P}^*)$.

- *Uniqueness of the solution of* (4.14).

  It suffices to note that $\ker(\mathcal{G}) \cap \ker(\mathcal{P}^*) = \mathrm{range}(\mathcal{P}) \cap \ker(\mathcal{P}^*) = \{0\}$.

Let $Z = C - \mathcal{A}^*\bar{y} \in \mathrm{range}(Q \cdot Q^\top)$. Since $\mathcal{A}^*$ is one-one, (4.16) implies that

$$\bar{Z} - \mathcal{A}^*y \in \mathbb{S}_+^n \iff \bar{Z} - \mathcal{A}^*y \in \mathbb{S}_+^n \text{ and } y = \mathcal{P}v.$$

Hence

$$\begin{aligned} v_{\mathrm{P}} &= b^\top\bar{y} + \sup\left\{b^\top(\mathcal{P}v) : QQ^\top\bar{Z}QQ^\top - \mathcal{A}^*\mathcal{P}v \succeq 0\right\} \\ &= b^\top\bar{y} + \sup\left\{b^\top(\mathcal{P}v) : Q^\top\bar{Z}Q - Q^\top(\mathcal{A}^*\mathcal{P}v)Q \succeq 0\right\}. \end{aligned}$$

This proves (4.13). Moreover, $v \in \mathbb{R}^{\bar{m}}$ is feasible for (4.13) if and only if $\bar{y} + \mathcal{P}v \in \mathcal{F}_{\mathrm{P}}^y$. Finally,

$$\bar{\mathcal{A}}^*v = 0 \iff \mathcal{A}^*\mathcal{P}v = QQ^\top(\mathcal{A}^*\mathcal{P}v)QQ^\top = 0 \iff v = 0,$$

i.e., $\bar{\mathcal{A}}^*$ is one-one. □

4.3.1 The finite number of iterations of the facial reduction

Since each iteration of the facial reduction (Algorithm 4.1) returns $\bar{\mathcal{L}}^{(k)} \subseteq \mathbb{S}^{n_k}$ with $n_k < n_{k-1}$ (taking $n_0 = n$), Algorithm 4.1 must terminate finitely. Indeed, the total number of facial reduction iterations cannot exceed $n - 1$.

For any fixed positive integer n, there do exist SDP instances that require $n - 1$ iterations of facial reductions to locate the minimal face of $\mathbb{S}_+^n$ containing the feasible region. Consider (P)

[1] $y \in \ker(\mathcal{G})$ if and only if $\mathcal{A}^*y = QQ^\top(\mathcal{A}^*y)QQ^\top$, if and only if $\mathcal{A}^*y = \mathcal{A}^*\mathcal{P}v$ for some $v \in \mathbb{R}^{\bar{m}}$, if and only if $y \in \mathrm{range}(\mathcal{P})$.

with the following input data [90]:

$$\begin{gathered} n \geq 3, \quad m = n; \quad b = e_2 \in \mathbb{R}^n, \quad C = 0, \\ A_1 = e_1 e_1^\top, \quad A_2 = e_1 e_2^\top + e_2 e_1^\top, \quad \text{and} \quad A_j = e_{j-1} e_{j-1}^\top + e_1 e_j^\top + e_j e_1^\top, \ \forall\, j = 3, \ldots, n. \end{gathered} \tag{4.17}$$

Then (P) with data given in (4.17) has the feasible region $\mathcal{F}_{\mathrm{P}}^Z = \{\mu e_1 e_1^\top : \mu \geq 0\}$, and requires $n-1$ iterations of facial reductions to find the minimal face $\operatorname{face}(\mathrm{P}) = \mathcal{F}_{\mathrm{P}}^Z$. To see this, note that $\mathcal{A}_C(D) = 0$ and $D \succeq 0$ if and only if $D = \gamma e_n e_n^\top$ for some $\gamma \geq 0$. Then the first step of the facial reduction gives

$$\mathcal{F}_{\mathrm{P}}^Z = \operatorname{range}(\mathcal{A}^*) \cap (\mathbb{S}_+^n \cap \{D\}^\perp) = \operatorname{span}(A_1, \ldots, A_{n-2}) \cap \begin{bmatrix} \mathbb{S}_+^{n-1} & 0 \\ 0 & 0 \end{bmatrix}.$$

Inductively, it would take $n-1$ facial reduction iterations to find the minimal face $\{\mu e_1 e_1^\top : \mu \geq 0\}$.

The number of facial reduction iterations required to find the minimal face of $\mathbb{S}_+^n$ containing the set $\bar{\mathcal{L}} \cap \mathbb{S}_+^n$, where $\bar{\mathcal{L}} \subseteq \mathbb{S}^n$ is a linear subspace, is also called the *degree of singularity* of $\bar{\mathcal{L}}$; see Definition 7.5.5.

4.4 Auxiliary problem

In this section, we consider one way of finding a nonzero element in $\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}}$ (defined in (4.1)). Determining whether $\mathcal{R}_{\mathrm{D}_{\mathrm{conic}}}$ contains a nonzero element is crucial for finding a smaller face containing the feasible slacks of ($\mathrm{P}_{\mathrm{conic}}$) (see (4.2) in Lemma 4.1.1), allowing for the regularization of ($\mathrm{P}_{\mathrm{conic}}$) by refining the feasible region (see (4.5)).

We consider the *auxiliary problem* for ($\mathrm{P}_{\mathrm{conic}}$):

$$\begin{aligned} v_{\mathrm{aux}} = \ & \inf_{\delta, D} \ \delta \\ & \text{s.t.} \ \ \|\mathcal{A}_C(D)\| \leq \delta, \\ & \qquad \langle E, D \rangle_{\mathbb{V}} = 1, \\ & \qquad D \in \mathcal{K}^*, \end{aligned} \tag{4.18}$$

where $\mathcal{A}_C$ is defined in (3.15), and $0 \neq E \in \operatorname{ri}(\mathcal{K})$ is arbitrary (though for specific choices of $\mathcal{K}$ we would use particular choices of E).

The auxiliary problem is a mixed conic program, involving the ordering cone $\mathcal{K}^* \times \mathcal{Q}^{m+2}$. The second order cone constraint $\|\mathcal{A}_C(D)\| \leq \delta$ together with the objective of minimizing δ aims

at finding some $D \in \mathcal{K}^*$ such that $\mathcal{A}_C(D) = 0$. The linear constraint $\langle E, D\rangle_{\mathbb{V}} = 1$ serves as a normalization constraint, forcing any feasible D to be nonzero. The dual of (4.18) is given by[2]

$$\begin{array}{ll} \sup\limits_{\beta,z,W} & \beta \\ \text{s.t.} & \mathcal{A}_C^* z + \beta E + W = 0, \\ & \|z\| \leq 1, \\ & W \in \mathcal{K}. \end{array} \tag{4.19}$$

4.4.1 Basic facts about the auxiliary problem

We first list some basic facts about (4.18) and its dual, provided that $\mathcal{K}$ is a proper cone.

Proposition 4.4.1. *Suppose that $\mathcal{K}$ is a proper cone and $E \in \operatorname{int}(\mathcal{K})$. Then the conic program* (4.18) *and its dual* (4.19) *both satisfy the Slater condition.*

Moreover, the optimal value v_{aux} of (4.18) *equals 0 if and only if $\mathcal{R}_{\text{D}_{\text{conic}}} \neq \{0\}$. In particular, if* ($\text{P}_{\text{conic}}$) *is feasible, then $v_{\text{aux}} = 0$ if and only if the Slater condition does not hold for* (P_{conic}).

If $v_{\text{aux}} = 0$, then $(0, D)$ is an optimal solution of (4.18) *if and only if $D \in \mathcal{R}_{\text{D}_{\text{conic}}}$.*

Proof. Pick any $D_0 \in \operatorname{int}(\mathcal{K}^*)$. Then $E \in \operatorname{int}(\mathcal{K})$ implies that $\langle E, D_0\rangle_{\mathbb{V}} > 0$ (see Lemma 3.3.2). Letting $D = \frac{1}{\langle E, D_0\rangle_{\mathbb{V}}} D_0$ and picking any $\delta > \|\mathcal{A}_C(D)\|$, (δ, D) is a Slater point of (4.18). On the other hand, $(-1, 0, E)$ is a Slater point of (4.19). This shows that both (4.18) and (4.19) satisfy the Slater condition. In particular, strong duality holds and (4.18) has an optimal solution.

Observe that $v_{\text{aux}} \geq 0$ and any feasible solution D of (4.18) has to be nonzero because of the constraint $\langle E, D\rangle_{\mathbb{V}} = 1$. If $0 \neq D_0 \in \mathcal{R}_{\text{D}_{\text{conic}}}$, then $\langle E, D_0\rangle_{\mathbb{V}} > 0$ and $D = \frac{1}{\langle E, D_0\rangle_{\mathbb{V}}} D_0 \in \mathcal{R}_{\text{D}_{\text{conic}}}$. Hence $(0, D)$ is feasible, implying that $v_{\text{aux}} = 0$. If $v_{\text{aux}} = 0$, then for any optimal solution $(0, \bar{D})$ of (4.18), we have $0 \neq \bar{D} \in \mathcal{R}_{\text{D}_{\text{conic}}}$. Therefore $v_{\text{aux}} = 0$ if and only if $\mathcal{R}_{\text{D}_{\text{conic}}} \neq \{0\}$. In particular, if ($\text{P}_{\text{conic}}$) is feasible, then by Theorem 3.3.10 $v_{\text{aux}} = 0$ if and only if the Slater condition does not hold for (P_{conic}). □

[2] The Lagrangian of (4.18) is

$$\begin{aligned} \text{L}(\delta, D; \alpha, z, \beta, W) &= \delta - (\alpha\delta + (\mathcal{A}_C(D))^\top z) + \beta(1 - \langle E, D\rangle_{\mathbb{V}}) - \langle D, W\rangle_{\mathbb{V}} \\ &= (1 - \alpha)\delta - \langle D, \mathcal{A}_C^* z + \beta E + W\rangle_{\mathbb{V}} + \beta. \end{aligned}$$

Remark. In the case of SDP, we take $E = \frac{1}{\sqrt{n}} I$.[3] If $v_{\text{aux}} > 0$, then for any $Z = C - \mathcal{A}^* y \in \mathbb{S}^n_+$, $\lambda_{\min}(Z) \leq v_{\text{aux}} \sqrt{1 + \|y\|^2}$. In fact, $Z \succeq \lambda_{\min}(Z) I$ and $\langle I, D \rangle = \sqrt{n}$ imply that

$$\sqrt{n}\, \lambda_{\min}(Z) \leq \langle D, Z \rangle = (\mathcal{A}_C(D))^\top \begin{pmatrix} -y \\ 1 \end{pmatrix} \leq \|\mathcal{A}_C(D)\| \sqrt{1 + \|y\|^2} = v_{\text{aux}} \sqrt{1 + \|y\|^2},$$

i.e.,

$$\lambda_{\min}(Z) \leq \frac{1}{\sqrt{n}} v_{\text{aux}} \sqrt{1 + \|y\|^2}. \tag{4.20}$$

This illustrates that when v_{aux} is close to zero, the smallest eigenvalue of any feasible slack has to be also close to zero in a relative sense.

We can generalize the result (4.20) in the remark to general conic programs (P_{conic}), to show that the optimal value v_{aux} of (4.18) is a measure of how close the Slater condition is to failing. While (4.20) concerns the smallest eigenvalue of the feasible slacks of (P_{conic}), the following result concerns how "close" the set of feasible slacks $\mathcal{F}^Z_{\text{P}_{\text{conic}}}$ is to being contained in a proper face of $\mathcal{K}$, in terms of the cosine of the angle between $\mathcal{F}^Z_{\text{P}_{\text{conic}}}$ and the face $\mathcal{K} \cap \{\bar{D}\}^\perp$, where $(\bar{\delta}, \bar{D})$ is any optimal solution of (4.18).

Theorem 4.4.2. *[27, Theorem 12.17] Assume that* (P_{conic}) *is feasible, that $\mathcal{K}$ is a proper cone and that $\mathcal{A}$ is onto. Let $E \in \text{int}(\mathcal{K})$ with $\|E\|_{\mathbb{V}} = 1$. If (δ, D) is a feasible solution of the auxiliary problem* (4.18)*, then either $\mathcal{F}^Z_{\text{P}_{\text{conic}}} = \{0\}$ or*

$$0 \leq \sup_{0 \neq Z = C - \mathcal{A}^* y \in \mathcal{K}} \frac{\langle D, Z \rangle_{\mathbb{V}}}{\|D\|_{\mathbb{V}} \|Z\|_{\mathbb{V}}} \leq \alpha(\mathcal{A}_C, \delta) := \begin{cases} \dfrac{\delta}{\sigma_{\min}(\mathcal{A}^*)} & \text{if } C \in \text{range}(\mathcal{A}^*), \\ \dfrac{\delta}{\sigma_{\min}(\mathcal{A}^*_C)} & \text{if } C \notin \text{range}(\mathcal{A}^*), \end{cases} \tag{4.21}$$

where $\sigma_{\min}(\mathcal{A}^*)$ is defined in (3.1) *and $\sigma_{\min}(\mathcal{A}^*_C)$ is similarly defined.*

Proof. If $\mathcal{F}^Z_{\text{P}_{\text{conic}}} \neq \{0\}$, then the optimization problem in (4.21) is feasible. Note also that

$$\langle E, D \rangle_{\mathbb{V}} = 1 = \|E\|_{\mathbb{V}} \implies \|D\|_{\mathbb{V}} \geq \frac{\langle E, D \rangle_{\mathbb{V}}}{\|E\|_{\mathbb{V}}} = 1.$$

If $C = \mathcal{A}^* y_C$ for some $y_C \in \mathbb{R}^m$, then for any $y \in \mathbb{R}^m$, $Z := C - \mathcal{A}^* y = \mathcal{A}^*(y_C - y)$ is nonzero if and only if $y \neq y_C$, and $\|Z\|_{\mathbb{V}} \geq \sigma_{\min}(\mathcal{A}^*) \|y - y_C\|$. Hence

$$\frac{\langle D, Z \rangle_{\mathbb{V}}}{\|D\|_{\mathbb{V}} \|Z\|_{\mathbb{V}}} \leq \frac{(\mathcal{A}(D))^\top (y_C - y)}{\sigma_{\min}(\mathcal{A}^*) \|y - y_C\|} \leq \frac{\|\mathcal{A}_C(D)\| \|y - y_C\|}{\sigma_{\min}(\mathcal{A}^*) \|y - y_C\|} \leq \frac{\delta}{\sigma_{\min}(\mathcal{A}^*)}.$$

[3] More generally, if $\mathcal{K}$ is a symmetric cone, then we would choose E to be a positive scalar multiple of the multiplicative identity.

If $C \notin \text{range}(\mathcal{A}^*)$, then $\sigma_{\min}(\mathcal{A}_C^*) > 0$. For any $y \in \mathbb{R}^m$, $Z := C - \mathcal{A}^* y = \mathcal{A}_C^* \begin{pmatrix} -y \\ 1 \end{pmatrix}$ satisfies $\|Z\| \geq \sigma_{\min}(\mathcal{A}_C^*)\sqrt{1+\|y\|^2}$, so

$$\frac{\langle D, Z\rangle_{\mathbb{V}}}{\|D\|_{\mathbb{V}}\|Z\|_{\mathbb{V}}} \leq \frac{(\mathcal{A}_C(D))^\top \begin{pmatrix} -y \\ 1 \end{pmatrix}}{\sigma_{\min}(\mathcal{A}_C^*)\sqrt{1+\|y\|^2}} \leq \frac{\delta}{\sigma_{\min}(\mathcal{A}_C^*)}.$$

This proves (4.21). □

In particular, if (δ^*, D^*) is an optimal solution of the auxiliary problem (4.18), then

$$0 \leq \sup_{0 \neq Z = C - \mathcal{A}^* y \in \mathcal{K}} \frac{\langle D^*, Z\rangle_{\mathbb{V}}}{\|D^*\|_{\mathbb{V}}\|Z\|_{\mathbb{V}}} \leq \alpha(\mathcal{A}_C) := \alpha(\mathcal{A}_C, v_{\text{aux}}),$$

and we recover the result that $v_{\text{aux}} = 0$ implies $\langle D^*, Z\rangle_{\mathbb{V}} = 0$ for all $Z \in \mathcal{F}^Z_{\text{P}_{\text{conic}}}$.

4.4.2 Strict complementarity of the auxiliary problem

Given that the Slater condition always holds for (4.18)-(4.19) when $\mathcal{K}$ is a proper cone, the necessary and sufficient optimality condition for (4.18)-(4.19) is given by

$$\begin{array}{llll} \mathcal{A}_C^* z + \beta E + W = 0, & \|z\| \leq 1, & W \in \mathcal{K}, & \text{(dual feasibility)} \\ \langle E, D\rangle_{\mathbb{V}} = 1, & \|\mathcal{A}_C(D)\| \leq \delta, & D \in \mathcal{K}^*, & \text{(primal feasibility)} \\ & \delta + (\mathcal{A}_C(D))^\top z = 0, & \langle D, W\rangle_{\mathbb{V}} = 0. & \text{(complementary slackness)} \end{array} \tag{4.22}$$

When the feasible program (P_{conic}) fails the Slater condition, the maximally complementary optimal solutions of (4.18)-(4.19) gives us information about whether the reduced problem (4.3) satisfies the Slater condition, i.e., whether face $(\text{P}) = \mathcal{K} \cap (\mathcal{R}_{\text{D}_{\text{conic}}})^\perp$.

Proposition 4.4.3. *[27, Theorem 12.28] Let $(0, D; 0, z, W)$ be a maximally complementary solution of* (4.18)-(4.19)*. Then the reduced program* (4.3) *satisfies the Slater condition if and only if $W \in \text{ri}(\mathcal{K} \cap \{D\}^\perp)$.*

Proof. Suppose that (4.3) satisfies the Slater condition, i.e., there exists $\hat{y} \in \mathbb{R}^m$ such that $C - \mathcal{A}^*\hat{y} \in \text{ri}(\mathcal{K} \cap \{D\}^\perp)$. Let

$$\bar{\beta} = 0, \quad \bar{z} = \frac{1}{\sqrt{1+\|\hat{y}\|^2}} \begin{pmatrix} \hat{y} \\ -1 \end{pmatrix}, \quad \bar{W} = -\mathcal{A}_C^* \bar{z} = \frac{1}{\sqrt{1+\|\hat{y}\|^2}}(C - \mathcal{A}^*\hat{y}).$$

Then $(\bar{\beta}, \bar{z}, \bar{W})$ is an optimal solution of (4.19) and $\bar{W} \in \text{ri}(\mathcal{K} \cap \{D\}^\perp)$. Hence the maximally complementary solution satisfies $W \in \text{ri}(\mathcal{K} \cap \{D\}^\perp)$ as well.

Conversely, if $W \in \operatorname{ri}(\mathcal{K} \cap \{D\}^\perp)$, then $\operatorname{range}(\mathcal{A}_C^*) \cap \operatorname{ri}(\mathcal{K} \cap \{D\}^\perp) \neq \emptyset$. Hence (4.3) satisfies the Slater condition by Lemma 3.3.6. □

In other words, the existence of a strictly complementary optimal solution of the auxiliary problem with optimal value zero implies that only one iteration of facial reduction is needed to find the minimal face.

In the special case of linear programs, we expect that at most one facial reduction iteration is required to arrive at the minimal face, since the corresponding auxiliary problem would be equivalent to an LP. In fact, Freund *et al.* [40] showed that the minimal face of

$$v_{\mathrm{P_{LP}}} = \max_y \left\{ b^\top y : c - A^\top y \geq 0 \right\} \tag{$\mathrm{P_{LP}}$}$$

can be identified by solving an auxiliary LP. We state the simpler version of the two results from [40].

Theorem 4.4.4. *[40, Proposition 1] Suppose that* ($\mathrm{P_{LP}}$) *is feasible, i.e.,* $\mathcal{F}_{\mathrm{LP}} := \left\{ y : c - A^\top \geq 0 \right\} \neq \varnothing$. *Consider the linear program*

$$\begin{array}{ll} \max\limits_{y,z,\alpha} & \bar{e}^\top z \\ \text{s.t.} & A^\top y + z - \alpha c \leq 0, \\ & 0 \leq z \leq \bar{e}, \\ & \alpha \geq 1, \end{array} \tag{4.23}$$

where $\bar{e}$ *is the vector of all ones of appropriate length. Then* (4.23) *is feasible and finite, and for any optimal solution* (y^*, z^*, α^*) *of* (4.23),

$$\left\{ i \in 1 : m : (c - A^\top y)_i = 0,\ \forall\, y \in \mathcal{F}_{\mathrm{LP}} \right\} = \{ i \in 1 : m : z_i^* = 0 \} \quad \textit{and} \quad \frac{1}{\alpha^*} y^* \in \operatorname{ri}(\mathcal{F}_{\mathrm{LP}}).$$

In other words, any optimal solution (y^*, z^*, α^*) of (4.23) determines the minimal face of $\mathbb{R}^n$ containing $\mathcal{F}_{\mathrm{LP}}$.

We apply Proposition 4.4.3 to prove a similar result. We show that linear programs only require at most one facial reduction iteration to arrive at the minimal face.

Corollary 4.4.5. *If the linear program* ($\mathrm{P_{LP}}$) *is feasible and does not satisfy the Slater condition, then the reduced program*

$$\max_y \left\{ b^\top y : c - A^\top y \in \mathbb{R}^n_+ \cap \{d\}^\perp \right\}, \tag{4.24}$$

where $d \in \operatorname{ri}\left(\left\{ g \in \mathbb{R}^n : Ag = 0,\ c^\top g = 0,\ g \geq 0 \right\} \right)$ *is arbitrary, satisfies the Slater condition.*

Proof. Consider the LP

$$\min\left\{0 : Ag = 0,\ c^\top g = 0,\ \bar{e}^\top g = 1,\ g \geq 0\right\}. \tag{4.25}$$

Since ($\mathrm{P_{LP}}$) fails the Slater condition, the LP (4.25) is feasible. The dual of (4.25) is

$$\max\left\{\alpha : \alpha\bar{e} + \gamma c + A^\top y \leq 0\right\} \tag{4.26}$$

and is also feasible. By the Goldman-Tucker theorem (Theorem 3.2.7), (4.25) and (4.26) have strictly complementary solutions, i.e., there exist $g \geq 0$ with $Ag = 0$, $c^\top g = 0$, $\bar{e}^\top g = 1$ and (γ, y) with $\gamma + \|y\|^2 < 1$ and $w := -\gamma c - A^\top y \geq 0$ such that $w + g > 0$. Then $(0, g; 0, (y; \gamma), w)$ is an optimal solution of (4.18)-(4.19) (with $\mathcal{K} = \mathbb{R}^n$). Then $d \in \mathrm{ri}\left(\left\{g \in \mathbb{R}^n : Ag = 0,\ c^\top g = 0,\ g \geq 0\right\}\right)$ implies that $(0, d; 0, (y; \gamma), w)$ is a maximally complementary solution of (4.18)-(4.19) with $w \in \mathrm{ri}(\mathbb{R}^n_+ \cap \{g\}^\perp) = \mathrm{ri}(\mathbb{R}^n_+ \cap \{d\}^\perp)$. Therefore the reduced program (4.24) satisfies the Slater condition. □

We remark that the single SOCP also only require at most one facial reduction iteration for finding the minimal face (see Theorem 4.2.1).

Part II

Numerical implementation of facial reduction on SDP

Chapter 5

Implementing facial reduction on SDP: numerical issues

From this chapter on, we will focus on semidefinite programs. The implementation of facial reduction on SDP is more challenging than on LP and SOCP, where at most one iteration of facial reduction is required[1]. Unlike LP and SOCP, SDP may require more than one iteration of facial reduction.

In this chapter, we study the implementation of one iteration of facial reduction on SDP and discuss the associated numerical issues. The main steps of one iteration of facial reduction are:

(1) finding an element $D \in \mathrm{ri}(\mathcal{R}_{\mathrm{D}})$;

(2) (if $0 \neq D \notin \mathbb{S}^n_{++}$) computing $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{span}(\{D\}^{\perp} \cap \mathbb{S}^n_+)$, and projecting the feasible slacks in $\{D\}^{\perp} \cap \mathbb{S}^n_+$ onto $\mathbb{S}^{\bar{n}}_+$ (where $\bar{n} = \dim(\ker(D))$).

Note that our facial reduction algorithm aims at finding not only the minimal face (by repeating step (1)), but also the minimal subspace (via step (2)). The projection in step (2) onto a smaller PSD cone and the use of a minimal subspace is essential for proper regularization; see [91].

One iteration of facial reduction on (P) is outlined in Algorithm 5.1. In practice, however, there are multiple occasions in Algorithm 5.1 where the computation is done only approximately:

- *Step (1)*: in practice, the auxiliary problem (4.18) is only solved approximately. In particular, we may arrive at a near optimal solution (δ^*, D^*) where $\delta^* \approx 0$.

[1] See Corollary 4.4.5 for the result on LP and Theorem 4.2.1 for the result on the SOCP ($\mathrm{P}_{\mathrm{SOCP}}$).

Algorithm 5.1: One simplified iteration of facial reduction algorithm on (P)

(1) **Solve the *auxiliary problem* for an optimal solution** (δ^*, D^*)**.**

If $\delta^* > 0$, then the Slater condition holds; **stop**.

If $\delta^* = 0$, then the Slater condition fails; **proceed to Step (2)**.

(2) **Compute the spectral decomposition of** D^***.**

If $D^* \succ 0$, then **stop**.

Otherwise, $D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P^\top \\ Q^\top \end{bmatrix}$, where $Q = \begin{bmatrix} P & Q \end{bmatrix}$ is orthogonal and $D_+ \succ 0$; **proceed to Step (3).**

(3) **Compute the subspace intersection.**

If $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top) = \{0\}$, then **stop**.

Otherwise, find a one-one linear map $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$ with

$$\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top) = \mathrm{range}(\mathcal{A}^*\mathcal{P});$$

proceed to Step (4).

(4) **Shift the objective**

Solve the linear equations (4.14) for the unique solution $\bar{y}$; **proceed to Step (5)**.

(5) **Project the problem data**

$\bar{\mathcal{A}}^*(\cdot) \leftarrow Q^\top (\mathcal{A}^*\mathcal{P}(\cdot))\, Q$;

$\bar{b} \leftarrow \mathcal{P}^* b$;

$\bar{C} \leftarrow Q^\top (C - \mathcal{A}^*\bar{y}) Q$.

% `Then` (4.13) `holds; see Item (2) of Theorem 4.3.2.`

- *Step (2)*: in practice, we would get

$$D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P^\top \\ Q^\top \end{bmatrix}, \quad \text{with } D_\epsilon \approx 0,$$

  and we round D_ϵ down to zero. The decision of which eigenvalues of D^* are zero affects our choice of Q, the nullspace of D^* which determines the smaller face of $\mathbb{S}^n_+$ containing $\mathcal{F}^Z_{\mathrm{P}}$. (In particular, Q is used in Step (5) for projecting the problem data.)

- *Step (3)*: as we will see in Section 5.3, the computation of subspace intersection will involve the decision of whether the cosine of a *principal angle* equals 1. Again, the cosines of the principal angles can be calculated approximately only.

- *Step (4)*: the equation (4.14a) may no longer have a solution. In particular, it may be impossible to decompose C into the sum $\mathcal{A}^*y + QWQ^\top$. (Does there at least exist some (y, W) such that $C \approx \mathcal{A}^*y + QWQ^\top$? Does replacing C by $QWQ^\top$ lead to big changes in (P), loosely speaking?)

The approximations listed above would affect the quality of the computed equivalent SDP (4.13). In this chapter, we study the effects of these approximations. We first address the issue of rounding small eigenvalues of D^* to zero in Section 5.1. Then we consider the issue of inaccuracy in solving the auxiliary problem in Section 5.2. (We already started the discussion on the auxiliary problem for general conic program in Section 4.4; we strengthen the results from Theorem 4.4.2 in the case of SDP.) In Section 5.3, we discuss the computation of the subspace intersection $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top)$. In Section 5.4, we discuss the solution of $C = \mathcal{A}^*y + QWQ^\top$, so that we can shift the objective and allow for the projection of problem data.

Most of the results in this chapter are from [27].

5.1 Numerical rank and dimension reduction

The results in, e.g., Theorem 4.3.2, assume that, given $0 \neq D^* \in \mathcal{R}_{\mathrm{D}}$ being not positive definite, we get the exact factorization $D^* = PD_+P^\top$. In practice, however, the spectral decomposition of D^* would be in the form

$$D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P^\top \\ Q^\top \end{bmatrix}, \quad \text{with} \quad D_\epsilon \approx 0.$$

We need to decide which of the eigenvalues are small enough and can be rounded down to zero. This is important for determining Q: rounding large eigenvalues to 0 gives a larger subspace

range(Q), which means that the computed face $Q\mathbb{S}^{\bar{n}}_{+}Q^\top$ containing the feasible slacks $\mathcal{F}^Z_{\rm P}$ of (P) would be too "large".

We can partition D^* using the idea of numerical rank. Suppose that $\lambda_1(D^*) \geq \lambda_2(D^*) \geq \cdots \geq \lambda_n(D^*) \geq 0$; the *numerical rank* of D^* with respect to a positive constant $\gamma > 0$ is defined to be the integer $\mathrm{rank}(D^*, \gamma)$ such that

$$\lambda_{\mathrm{rank}(D^*,\gamma)}(D^*) > \gamma \geq \lambda_{\mathrm{rank}(D^*,\gamma)+1}(D^*).$$

Fix some $\varepsilon \in (0,1)$. For the given D^*, let $\gamma = \frac{\varepsilon\|D^*\|}{\sqrt{n}}$. Take $r \leftarrow \mathrm{rank}(D^*, \gamma) = \mathrm{rank}\left(D^*, \frac{\varepsilon\|D^*\|}{\sqrt{n}}\right)$,

$$D_+ \leftarrow \mathrm{Diag}\left(\lambda_1(D^*), \ldots, \lambda_r(D^*)\right), \quad D_\epsilon \leftarrow \mathrm{Diag}\left(\lambda_{r+1}(D^*), \ldots, \lambda_n(D^*)\right),$$

and partition the matrix of eigenvectors $\begin{bmatrix} P & Q \end{bmatrix}$ accordingly. Then

$$\lambda_{\min}(D_+) > \frac{\varepsilon\|D^*\|}{\sqrt{n}} \geq \lambda_{\max}(D_\epsilon) \implies \|D_\epsilon\| \leq \varepsilon\|D^*\|,$$

and

$$\frac{\|D_\epsilon\|^2}{\|D_+\|^2} = \frac{\|D_\epsilon\|^2}{\|D^*\|^2 - \|D_\epsilon\|^2} \leq \frac{\varepsilon^2\|D^*\|^2}{(1-\varepsilon^2)\|D^*\|^2} = \frac{1}{\varepsilon^{-2}-1} \approx 0,$$

i.e., $\|D_\epsilon\|$ is negligible comparing with $\|D_+\|$.

5.2 Auxiliary problem for SDP: numerical aspects

Recall the auxiliary problem for the SDP (P):

$$\begin{array}{rll} v_{\rm aux} = & \inf\limits_{\delta, D} & \delta \\ & \text{s.t.} & \|\mathcal{A}_C(D)\| \leq \delta, \\ & & \langle \frac{1}{\sqrt{n}} I, D\rangle = 1, \\ & & D \succeq 0. \end{array} \tag{5.1}$$

(We replace E in (4.18) by $\frac{1}{\sqrt{n}}I$.) Note that any feasible (δ, D) satisfies $1 \leq \|D\| \leq \sqrt{n}$.

Suppose we have a computed optimal solution (δ, D) of the auxiliary problem (5.1). For simplicity, we will assume that (δ, D) is at least feasible for (5.1). The inexactness of the solution introduces numerical errors in formulating the reduced problem (4.13). The main issue is: is (4.13) "far" from (P)? To formalize the discussion, in the following sections, we address a few numerical aspects around the use of an approximately optimal solution of the auxiliary problem to get (4.13).

- If $\delta \approx 0$, can we expect $\operatorname{dist}(Z, \mathcal{F}_{\mathrm{P}}^Z \cap \{D\}^\perp) \approx 0$ for any feasible $Z \in \mathcal{F}_{\mathrm{P}}^Z$? This will be discussed in Section 5.2.1.
- Can we use (δ, D) to formulate a theoretically equivalent problem of (P)? In fact, we can always rotate the feasible slacks, and a *rank-revealing rotation* can show that all feasible slacks lie in a somewhat "flat" cone; see Section 5.2.2.
- Can we avoid solving the auxiliary problem? In some cases, we can use a simple heuristic to find a nonzero element in $\mathcal{R}_{\mathrm{D}}$. In general, we use a preprocessing technique to reduce the problem size of (5.1); see Section 5.2.3.

We emphasize that in this section, though all the results are stated in terms of a feasible solution (δ, D) of the auxiliary problem (5.1), what we really are interested in is the situation when $\delta \approx 0$. In particular, $\alpha(\mathcal{A}_C, \delta)$ defined in (4.21) is also approximately zero.

5.2.1 Distance between $\mathcal{F}_{\mathrm{P}}^Z$ and the computed face $\mathcal{F}_{\mathrm{P}}^Z \cap \{D\}^\perp$

Suppose that $v_{\mathrm{aux}} \approx 0$ and we decide that (P) fails the Slater condition, i.e., that $\mathcal{F}_{\mathrm{P}}^Z \subseteq \mathbb{S}_+^n \cap \{D\}^\perp \lhd \mathbb{S}_+^n$, where (δ, D) is a computed optimal solution of (5.1). The computed smaller face $\mathbb{S}_+^n \cap \{D\}^\perp$ is either $\{0\}$ or $Q\mathbb{S}_+^{\bar{n}}Q^\top$, where $\operatorname{range}(Q) = \ker(D^*)$ and Q is of full column rank. Then the iteration of facial reduction essentially projects $\mathcal{F}_{\mathrm{P}}^Z$ onto either $\{0\}$ or $Q\mathbb{S}_+^{\bar{n}}Q^\top$. Naturally, we expect that at the very least, any arbitrary $Z \in \mathcal{F}_{\mathrm{P}}^Z$ is not far from the proper face $\mathbb{S}_+^n \cap \{D\}^\perp$ we wish to project it onto. Proposition 5.2.1 considers the case where the computed face equals $\{0\}$: we show that $\|Z\|$ is small in a relative sense. Proposition 5.2.2 considers the case where the computed face equals $Q\mathbb{S}_+^{\bar{n}}Q^\top$: we show that $\operatorname{dist}(Z, Q\mathbb{S}_+^{\bar{n}}Q^\top)$ is also relatively small.

We first show that the norm of any $Z \in \mathcal{F}_{\mathrm{P}}^Z$ is close to zero if we have a feasible solution (δ, D) with $\delta \approx 0$ and $D \succ 0$.

Proposition 5.2.1. *Let (δ, D) be a feasible solution of (5.1). If $\lambda_{\min}(D) > 0$, then*

$$\|Z\| \leq \frac{\delta}{\lambda_{\min}(D)}(1 + \|y\|^2)^{1/2} \tag{5.2}$$

for all $Z = C - \mathcal{A}^* y \succeq 0$.

Proof. The inequality (5.2) holds because

$$\lambda_{\min}(D)\|Z\| \leq \lambda_{\min}(D)\langle I, Z\rangle \leq \langle D, Z\rangle = (\mathcal{A}_C(D))^\top \begin{pmatrix} -y \\ 1 \end{pmatrix} \leq \delta \left\| \begin{pmatrix} -y \\ 1 \end{pmatrix} \right\|.$$

□

In particular, if $\delta = 0$ and $D \succ 0$, then (5.2) implies that $\mathcal{F}_{\mathrm{P}}^{Z} = \{0\}$. Now we consider the second case, where D is not positive definite.

Proposition 5.2.2. *[27, Proposition 12.18] Let (δ, D) denote a feasible solution of (5.1), where*

$$D = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P^\top \\ Q^\top \end{bmatrix}, \quad \begin{bmatrix} P & Q \end{bmatrix} \text{ orthogonal}, \quad D_+ \in \mathbb{S}_+^{n-\bar{n}} \text{ and } 0 < \bar{n} < n.$$

For any $0 \neq Z = C - \mathcal{A}^* y \succeq 0$, the angle between Z and its projection Z_Q onto $Q\mathbb{S}_+^{\bar{n}} Q^\top$ is small in the sense that

$$\cos\theta_{Z,Z_Q} := \frac{\langle Z, Z_Q\rangle}{\|Z\|\|Z_Q\|} = \frac{\|Q^\top Z Q\|}{\|Z\|} \geq 1 - \frac{\alpha(\mathcal{A}_C, \delta)\|D\|}{\lambda_{\min}(D_+)}, \tag{5.3}$$

where $\alpha(\mathcal{A}_C, \delta)$ is defined in (4.21). Moreover,

$$\operatorname{dist}(Z,\, Q\mathbb{S}_+^{\bar{n}} Q^\top) \leq \sqrt{2}\|Z\| \left(\frac{\alpha(\mathcal{A}_C, \delta)\|D\|}{\lambda_{\min}(D_+)}\right)^{1/2}. \tag{5.4}$$

Proof. Since $Z \succeq 0$, we have $QQ^\top ZQQ^\top = \arg\min_{W \in \operatorname{range}(Q\cdot Q^\top)} \|Z - W\| = \arg\min_{W \in Q\mathbb{S}_+^{\bar{n}} Q^\top} \|Z - W\|$. Since $D = PD_+P^\top + QD_\epsilon Q^\top$, by Theorem 4.4.2 Z satisfies

$$\langle PD_+P^\top, Z\rangle \leq \langle PD_+P^\top, Z\rangle + \langle QD_\epsilon Q^\top, Z\rangle = \langle D, Z\rangle \leq (\alpha(\mathcal{A}_C, \delta)\|D\|)\,\|Z\|.$$

Therefore

$$\frac{\langle Z, Z_Q\rangle}{\|Z\|\|Z_Q\|} = \frac{\langle Z, QQ^\top ZQQ\rangle}{\|Z\|\|Q^\top ZQ\|} = \frac{\|Q^\top ZQ\|}{\|Z\|} \geq \gamma, \tag{5.5}$$

where

$$\gamma := \min_{W \neq 0} \left\{ \frac{\|Q^\top W Q\|}{\|W\|} : \left\langle PD_+P^\top, \frac{W}{\|W\|} \right\rangle \leq \alpha(\mathcal{A}_C, \delta)\|D\|, \;\; W \succeq 0 \right\} \tag{5.6}$$

$$= \min_{W} \Big\{ \|Q^\top W Q\| \;:\; \langle PD_+P^\top, W\rangle \leq \alpha(\mathcal{A}_C, \delta)\|D\|, \;\; \|W\| = 1, \;\; W \succeq 0 \Big\}. \tag{5.7}$$

In the following, we compute the optimal value γ of (5.6). Using the orthogonal rotation

$$W = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^\top & S_{22} \end{bmatrix} \begin{bmatrix} P^\top \\ Q^\top \end{bmatrix} = PS_{11}P^\top + PS_{12}Q^\top + QS_{12}^\top P^\top + QS_{22}Q^\top,$$

we can rewrite (5.7) as

$$\begin{aligned} \gamma = \min_{S} \;\; & \|S_{22}\| \\ \text{s.t.} \;\; & \langle D_+, S_{11}\rangle \leq \alpha(\mathcal{A}_C, \delta)\|D\|, \\ & \|S_{11}\|^2 + 2\|S_{12}\|^2 + \|S_{22}\|^2 = 1, \\ & S = \begin{bmatrix} S_{11} & S_{12} \\ S_{12}^\top & S_{22} \end{bmatrix} \succeq 0. \end{aligned} \tag{5.8}$$

Note that for any $\begin{bmatrix} S_{11} & S_{12} \\ S_{12}^\top & S_{22} \end{bmatrix} \succeq 0$, we have

$$\|S_{12}\|^2 \le \|S_{11}\|\|S_{22}\| \implies (\|S_{11}\| + \|S_{22}\|)^2 \ge \|S_{11}\|^2 + 2\|S_{12}\|^2 + \|S_{22}\|^2.$$

Therefore $\gamma \ge \bar{\gamma}$, where

$$\begin{aligned} \bar{\gamma} &:= \min_{S_{11},S_{22}} \Big\{ \|S_{22}\| : \langle D_+, S_{11}\rangle \le \alpha(\mathcal{A}_C,\delta)\|D\|, \quad \|S_{11}\| + \|S_{22}\| \ge 1, \quad S_{11} \succeq 0, \quad S_{22} \succeq 0 \Big\} \\ &\ge \min_{S_{22}} \Big\{ 1 - \|S_{22}\| : \langle D_+, S_{22}\rangle \le \alpha(\mathcal{A}_C,\delta)\|D\|, \quad S_{22} \in \mathbb{S}_+^{n-\bar{n}} \Big\} \qquad (5.9) \\ &\ge 1 - \max_{S_{22}} \Big\{ \|S_{22}\| : \lambda_{\min}(D_+)\|S_{22}\| \le \alpha(\mathcal{A}_C,\delta)\|D\|, \quad S_{22} \in \mathbb{S}_+^{n-\bar{n}} \Big\} \\ &= 1 - \frac{\alpha(\mathcal{A}_C,\delta)\|D\|}{\lambda_{\min}(D_+)}. \end{aligned}$$

Let u be a normalized eigenvector of D_+ corresponding to its smallest eigenvalue $\lambda_{\min}(D_+)$. Then

$$\bar{S}^* := \frac{\alpha(\mathcal{A}_C,\delta)\|D\|}{\lambda_{\min}(D_+)}\, uu^\top \succeq 0 \quad \text{satisfies} \quad \langle D_+, \bar{S}^*\rangle = \alpha(\mathcal{A}_C,\delta)\|D\|, \quad \|\bar{S}^*\| = \frac{\alpha(\mathcal{A}_C,\delta)\|D\|}{\lambda_{\min}(D_+)},$$

i.e., $\bar{S}^*$ is feasible for (5.9), and $\|\bar{S}^*\| = \frac{\alpha(\mathcal{A}_C,\delta)\|D\|}{\lambda_{\min}(D_+)}$ implies that $\bar{S}^*$ is indeed an optimal solution of (5.9). In particular, $\bar{\gamma} = \frac{\alpha(\mathcal{A}_C,\delta)\|D\|}{\lambda_{\min}(D_+)}$.

Let $\beta := \min\Big\{1, \frac{\alpha(\mathcal{A}_C,\delta)\|D\|}{\lambda_{\min}(D_+)}\Big\}$; then $\gamma \ge 1-\beta$. Let $v \in \mathbb{R}^{\bar{n}}$ be an arbitrary unit-norm vector, and

$$S^* = \begin{bmatrix} S_{11}^* & S_{12}^* \\ (S_{12}^*)^\top & S_{22}^* \end{bmatrix} := \begin{pmatrix} \sqrt{\beta}u \\ \sqrt{1-\beta}\, v \end{pmatrix} \begin{pmatrix} \sqrt{\beta}u \\ \sqrt{1-\beta}\, v \end{pmatrix}^\top = \begin{bmatrix} \beta uu^\top & \sqrt{\beta(1-\beta)}uv^\top \\ \sqrt{\beta(1-\beta)}vu^\top & (1-\beta)vv^\top \end{bmatrix}.$$

Then $S^* \succeq 0$ is feasible for (5.8):

- $\langle D_+, S_{11}^*\rangle = \beta\lambda_{\min}(D_+) \le \frac{\alpha(\mathcal{A}_C,\delta)\|D\|}{\lambda_{\min}(D_+)}$,
- $\|S_{11}^*\|^2 + 2\|S_{12}^*\|^2 + \|S_{22}^*\|^2 = \beta^2 + 2\beta(1-\beta) + (1-\beta)^2 = 1$.

Also, $\gamma \le \|S_{22}^*\| = 1-\beta \le \gamma$, so S^* is an optimal solution of (5.8) and $\gamma = 1-\beta$. From (5.5):

$$\frac{\|Q^\top ZQ\|}{\|Z\|} \ge 1-\beta \ge 1 - \frac{\alpha(\mathcal{A}_C,\delta)\|D^*\|}{\lambda_{\min}(D_+)}.$$

Therefore (5.3) holds. For (5.4),

$$\begin{aligned}
\operatorname{dist}(Z,\, Q\mathbb{S}^{\bar{n}}_{+}Q^\top) &= \|Z - QQ^\top ZQQ\| = \left(\|Z\|^2 - \|Q^\top ZQ\|^2\right)^{1/2} \\
&= \|Z\| \left(1 - \frac{\|Q^\top ZQ\|^2}{\|Z\|^2}\right)^{1/2} \\
&\le \|Z\| \left(1 - (1-\beta)^2\right)^{1/2} \\
&\le \|Z\| \left(2\beta\right)^{1/2} \\
&\le \sqrt{2}\, \|Z\| \left(\frac{\alpha(\mathcal{A}_C, \delta)\|D^*\|}{\lambda_{\min}(D_+)}\right)^{1/2}.
\end{aligned}$$

□

Remark. If $\delta = 0$, then (5.4) implies that $\mathcal{F}^Z_{\mathrm{P}} \subseteq Q\mathbb{S}^{\bar{n}}_{+}Q^\top$.

Even though D_ϵ does not appear in the bound (5.4), the inequality (5.4) still makes sense: if D_ϵ is not close to zero, then the face $Q\mathbb{S}^{\bar{n}}_{+}Q^\top$ would be "too big" (but still containing a proper face that contains $\mathcal{F}^Z_{\mathrm{P}}$).

5.2.2 Rank-revealing rotation and equivalent problems

We saw in Theorem 4.3.2 that if we can solve the auxiliary problem exactly and obtain $0 \neq D \in \mathcal{R}_{\mathrm{D}}$ with $D \not\succ 0$, then we may either get the single feasible solution of (P) or rewrite (P) as a smaller equivalent problem (4.13).

Given a computed optimal solution (δ, D) of (5.1) with $\delta \approx 0$, while the reduced problem that we will arrive at is not going to be exactly equivalent to (P), the computed solution (δ, D) can still be used for rotating (the feasible slacks of) the SDP so that the feasible slacks are roughly of the form $\begin{bmatrix} * & 0 \\ 0 & 0 \end{bmatrix}$.

Proposition 5.2.3. *Let (δ, D) be a feasible solution of (5.1), and suppose that*

$$D = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P^\top \\ Q^\top \end{bmatrix}, \quad U = \begin{bmatrix} P & Q \end{bmatrix} \text{ orthogonal}, \quad \lambda_{\min}(D_+) > 0.$$

Then (P) is equivalent to the following problems:

$$\sup_y \left\{ b^\top y : U^\top ZU \in T_\beta,\ Z = C - \mathcal{A}^* y \right\}, \tag{5.10}$$

where $\beta := \alpha(\mathcal{A}_C, \delta)\frac{\|D\|}{\lambda_{\min}(D_+)}$ and

$$T_\beta := \left\{ Z = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix} \in \mathbb{S}^n_+ : \operatorname{tr}(A) \le \beta \operatorname{tr}(Z), \quad C \in \mathbb{S}^{\bar{n}} \right\},$$

which is a proper face of $\mathbb{S}^n_+$ *if* $\beta = 0$*, or a proper cone otherwise.*

Proof. If $Z = C - \mathcal{A}^* y \succeq 0$, then by (4.21),

$$\begin{aligned}\alpha(\mathcal{A}_C, \delta)\|D\|\|Z\| \geq \langle D, Z\rangle &= \langle PD_+P^\top + QD_\epsilon Q^\top, Z\rangle \\ &\geq \langle PD_+P^\top, Z\rangle \geq \lambda_{\min}(D_+)\langle I, P^\top ZP\rangle.\end{aligned}$$

Hence $\operatorname{tr}(P^\top ZP) \leq \beta\|Z\| \leq \beta \operatorname{tr}(Z) = \beta \operatorname{tr}(U^\top ZU)$, i.e., $U^\top ZU = \begin{bmatrix} P^\top ZP & P^\top ZQ \\ Q^\top ZP & Q^\top ZQ \end{bmatrix} \in T_\beta$. Therefore

$$Z = C - \mathcal{A}^* y \in \mathbb{S}^n_+ \iff Z = C - \mathcal{A}^* y \in T_\beta,$$

and (P) is equivalent to (5.10).

The set T_β is the intersection of $\mathbb{S}^n_+$ with a closed half space, so T_β is a nonempty pointed closed convex cone. If $\beta = 0$, then $T_\beta = \mathbb{S}^n_+ \cap \left\{ \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \right\}^\perp$ is a proper face of $\mathbb{S}^n_+$. If $\beta > 0$, then $\begin{bmatrix} \frac{\beta}{2} I & 0 \\ 0 & I \end{bmatrix} \in \operatorname{int}(T_\beta)$, implying that T_β is a proper cone. □

5.2.3 Preprocessing the auxiliary problem and a heuristic for finding $0 \neq D \in \mathcal{R}_{\mathrm{D}}$

In this section, we mention a preprocessing procedure for solving (5.1). The preprocessing deals with two scenarios.

(1) Suppose that A_i is positive semidefinite for some $i \in 0 : m$ (where $A_0 := C$). Let $\tilde{U} \in \mathbb{R}^{n \times r}$ be of full column rank and satisfy $\operatorname{range}(\tilde{U}) = \ker(A_i)$. Then $D \in \mathcal{R}_{\mathrm{D}}$ implies that $D \in \mathbb{S}^n_+ \cap \{A_i\}^\perp = \tilde{U}\mathbb{S}^r_+\tilde{U}^\top$. Therefore

$$D \in \mathcal{R}_{\mathrm{D}} \iff D = \tilde{U}\tilde{D}\tilde{U}^\top,\ \langle \tilde{D}, \tilde{U}^\top A_j \tilde{U}\rangle = 0,\ \forall\, i \in 0 : m.$$

So we replace A_j by $\tilde{U}^\top A_j \tilde{U}$ before solving the auxiliary problem.

(2) Suppose that $0 \neq v \in \mathbb{R}^n$ satisfies $A_j v = 0$ for all $j \in 0 : m$. Then $\langle A_j, vv^\top\rangle = 0$ for all $j \in 0 : m$, i.e., $vv^\top \in \mathcal{R}_{\mathrm{D}}$.

While scenarios (1) and (2) occur with probability zero for any randomly generated SDP (recalling that the set of all singular matrices in $\mathbb{R}^{n \times n}$ is of Lebesgue measure zero), in practice scenarios (1) and (2) occur more frequently than the theory suggests. SDP relaxations arising from

many applications (especially combinatorial optimization problems) often have sparse constraint matrices that fall in one of the two possible scenarios; see Chapter 8 for some examples.

Both determining whether a matrix is positive semidefinite (in Case (1)) and determining whether $\bigcap_{j=0}^{m} \ker(A_j) \neq 0$ (in Case (2)) can be done by performing spectral decompositions on the matrices $C, A_1, \ldots, A_m$. Therefore the two scenarios offer the possibility of efficient preprocessing, stated in Algorithm 5.2 on Page 74.

Algorithm 5.2: Preprocessing for the auxiliary problem (5.1)

---

1 Input($A_0 := C,\ A_1, \ldots, A_m \in \mathbb{S}^n$)

(1) Any $A_i \succ 0$?

if *one of* A_i *(*$i \in 0:m$*) is definite* **then**
stop; (P) satisfies the Slater condition;

endif

(2) Any $A_i \succeq 0$?

while *one of* $A_i = \begin{bmatrix} U & \tilde{U} \end{bmatrix} \begin{bmatrix} \tilde{G} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} U^\top \\ \tilde{U}^\top \end{bmatrix}$ *with* $\tilde{G}$ *being definite (*$i \in 0:m$*)* **do**
$A_j \leftarrow \tilde{U}^\top A_j \tilde{U}$ for all $j \neq i$ and remove A_i;

endw

(3) $\bigcap_{j=0}^{m} \ker(A_j) \neq \{0\}$?

if $\exists\, V \in \mathbb{R}^{n \times k}$ *such that* $\|V\|^2 = \sqrt{n}$ *and* $A_j V = 0$ *for all* $j \in 0:m$ **then**
$\langle I, VV^\top \rangle = 1$ and $\langle A_j, VV^\top \rangle = 0$ for all $j \in 1:m$;

therefore $(0, VV^\top)$ is an optimal solution of the auxiliary problem

$$\min_{\delta, D} \left\{ \delta : \left(\sum_{j=0}^{m} \langle A_j, D \rangle^2 \right)^{1/2} \leq \delta,\ \langle I, D \rangle = \sqrt{n},\ D \succeq 0 \right\}. \tag{5.11}$$

else
use an SDP solver to solve (5.11).

endif

5.3 Subspace intersection

To compute $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top)$, we can use the principal angles between the subspaces $\mathrm{range}(\mathcal{A}^*)$ and $\mathrm{range}(Q \cdot Q^\top)$:

Theorem 5.3.1. *[47, Theorem 6.4.2] Let $Q \in \mathbb{R}^{n\times\bar{n}}$ be of full column rank and let $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ be an onto linear map. Let $r := \min\{\frac{1}{2}\bar{n}(\bar{n}+1), m\}$. Then there exist $U_1^{\mathrm{sp}}, \ldots, U_r^{\mathrm{sp}} \in \mathrm{range}(Q \cdot Q^\top), V_1^{\mathrm{sp}}, \ldots, V_r^{\mathrm{sp}} \in \mathrm{range}(\mathcal{A}^*)$ such that*

$$\begin{aligned}
\sigma_1^{\mathrm{sp}} &:= \langle U_1^{\mathrm{sp}}, V_1^{\mathrm{sp}}\rangle \\
&= \max_{U,V} \left\{ \langle U, V\rangle : \|U\| = 1 = \|V\|,\ U \in \mathrm{range}(Q \cdot Q^\top),\ V \in \mathrm{range}(\mathcal{A}^*) \right\}, \\
\sigma_k^{\mathrm{sp}} &:= \langle U_k^{\mathrm{sp}}, V_k^{\mathrm{sp}}\rangle \\
&= \max_{U,V} \Big\{ \langle U, V\rangle : \|U\| = 1 = \|V\|,\ U \in \mathrm{range}(Q \cdot Q^\top),\ V \in \mathrm{range}(\mathcal{A}^*), \\
&\qquad\qquad \langle U, U_i^{\mathrm{sp}}\rangle = 0 = \langle V, V_i^{\mathrm{sp}}\rangle,\ \forall\, i \in 1 : (k-1) \Big\},
\end{aligned} \tag{5.12}$$

for $k \in 2 : r$. Moreover, $1 \geq \sigma_1^{\mathrm{sp}} \geq \sigma_2^{\mathrm{sp}} \geq \cdots \geq \sigma_r^{\mathrm{sp}} \geq 0$. If $\sigma_1^{\mathrm{sp}} < 1$, then $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top) = \{0\}$. If $1 = \sigma_1^{\mathrm{sp}} = \cdots = \sigma_{\bar{m}}^{\mathrm{sp}} > \sigma_{\bar{m}+1}^{\mathrm{sp}} \geq \cdots \geq \sigma_r^{\mathrm{sp}} \geq 0$, then

$$\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top) = \mathrm{span}(U_1^{\mathrm{sp}}, U_2^{\mathrm{sp}}, \ldots, U_{\bar{m}}^{\mathrm{sp}}) = \mathrm{span}(V_1^{\mathrm{sp}}, V_2^{\mathrm{sp}}, \ldots, V_{\bar{m}}^{\mathrm{sp}}) = \mathrm{range}(\mathcal{A}^*\mathcal{P}),$$

where $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$ is the one-one linear map defined by $\mathcal{P}v := \sum_{i=1}^{\bar{m}} v_i (\mathcal{A}^*)^\dagger(V_i^{\mathrm{sp}})$.

We remark that the σ_i^{sp}'s, U_i^{sp}'s and V_i^{sp}'s in Theorem 5.3.1 can be computed using singular value decomposition, as in Algorithm 5.3 derived from [47, Algorithm 6.4.3].

In practice, we do not get $\sigma_i^{\mathrm{sp}} = 1$ exactly (for $i \in 1 : \bar{m}$). For a fixed tolerance $\varepsilon^{\mathrm{sp}} \in [0, 1)$, suppose that

$$1 \geq \sigma_1^{\mathrm{sp}} \geq \cdots \geq \sigma_{\bar{m}}^{\mathrm{sp}} \geq 1 - \varepsilon^{\mathrm{sp}} > \sigma_{\bar{m}+1}^{\mathrm{sp}} \geq \cdots \geq \sigma_r^{\mathrm{sp}} \geq 0. \tag{5.13}$$

We claim that the approximation

$$\mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\mathcal{A}^*) \approx \mathrm{span}(U_1^{\mathrm{sp}}, \ldots, U_{\bar{m}}^{\mathrm{sp}}) \approx \mathrm{span}(V_1^{\mathrm{sp}}, \ldots, V_{\bar{m}}^{\mathrm{sp}}) \tag{5.14}$$

is "good enough" if $\varepsilon^{\mathrm{sp}} \approx 0$, in the sense that for $i \in 1 : \bar{m}$, $U_i^{\mathrm{sp}} \in \mathrm{range}(Q \cdot Q^\top)$ satisfies

$$\mathrm{dist}(U_i^{\mathrm{sp}}, \mathrm{range}(\mathcal{A}^*)) \leq \|U_i^{\mathrm{sp}} - V_i^{\mathrm{sp}}\| = \sqrt{2 - 2\langle U_i^{\mathrm{sp}}, V_i^{\mathrm{sp}}\rangle} = \sqrt{2(1 - \sigma_i^{\mathrm{sp}})} \leq \sqrt{2\varepsilon^{\mathrm{sp}}},$$

and $V_i^{\mathrm{sp}} \in \mathrm{range}(\mathcal{A}^*)$ satisfies $\mathrm{dist}(V_i, \mathrm{range}(Q \cdot Q^\top)) \leq 2\varepsilon^{\mathrm{sp}}$ and

$$\|Q^\top V_i^{\mathrm{sp}} Q\| = \|QQ^\top V_i^{\mathrm{sp}} QQ^\top\| \|U_i^{\mathrm{sp}}\| \geq \langle Q^\top V_i^{\mathrm{sp}} Q, Q^\top U_i^{\mathrm{sp}} Q\rangle = \langle V_i^{\mathrm{sp}}, U_i^{\mathrm{sp}}\rangle = \sigma_i^{\mathrm{sp}} \geq 1 - \varepsilon^{\mathrm{sp}}.$$

Algorithm 5.3: Computing the subspace intersection $\mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\mathcal{A}^*)$

(1) Compute an orthonormal basis $\{V_1, V_2, \ldots, V_m\} \subset \mathbb{S}^n$ of $\mathrm{range}(\mathcal{A}^*)$ (using, e.g., QR-decomposition);

(2) form the matrix representation Ω of span $\left\{Q^\top V_1 Q, Q^\top V_2 Q, \ldots, Q^\top V_m Q\right\}$, i.e.,

$$\Omega \leftarrow \Big[\mathrm{svec}(Q^\top V_1 Q) \quad \mathrm{svec}(Q^\top V_2 Q) \quad \cdots \quad \mathrm{svec}(Q^\top V_m Q)\Big] \in \mathbb{R}^{\frac{1}{2}\bar{n}(\bar{n}+1)\times m},$$

where $\mathrm{svec} : \mathbb{S}^n \to \mathbb{R}^{\frac{1}{2}n(n+1)}$ is the invertible linear operator that satisfies $(\mathrm{svec}(X))^\top (\mathrm{svec}(Y)) = \langle X, Y\rangle$ for all $X, Y \in \mathbb{S}^n$, and $\mathrm{sMat} := \mathrm{svec}^{-1}$;

(3) compute the SVD of $\Omega = U^\Omega \Sigma^\Omega (V^\Omega)^\top$,

with $\Sigma = \mathrm{Diag}(\sigma_1^{\mathrm{sp}}, \sigma_2^{\mathrm{sp}}, \ldots)$ and $\sigma_1^{\mathrm{sp}} \geq \sigma_2^{\mathrm{sp}} \geq \cdots$;

(4) $U_i^{\mathrm{sp}} \leftarrow Q\,\mathrm{sMat}(U^\Omega_{:,i})Q^\top$;

$V_i^{\mathrm{sp}} \leftarrow \sum_{k=1}^m V^\Omega_{k,i} V_k$;

(5) if $\sigma_1^{\mathrm{sp}} < 1$, then $\mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\mathcal{A}^*) = \{0\}$;

otherwise $\mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\mathcal{A}^*) = \mathrm{span}(U_1^{\mathrm{sp}}, \ldots, U_{\bar{m}}^{\mathrm{sp}}) = \mathrm{span}(V_1^{\mathrm{sp}}, \ldots, V_{\bar{m}}^{\mathrm{sp}})$,

where $\bar{m}$ satisfies $\sigma_{\bar{m}}^{\mathrm{sp}} = 1 > \sigma_{\bar{m}+1}^{\mathrm{sp}}$.

Proposition 5.3.2. *Let $Q \in \mathbb{R}^{n\times\bar{n}}$ be of full column rank and let $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ be an onto linear map. Let $r := \min\{\frac{1}{2}\bar{n}(\bar{n}+1), m\}$ and $\varepsilon^{\rm sp} > 0$. Let $\sigma_i^{\rm sp}, U_i^{\rm sp}, V_i^{\rm sp}$ for $i \in 1 : r$ be defined as in (5.12). Define $\breve{\mathcal{A}} : \mathbb{S}^n \to \mathbb{R}^m$ by*

$$\breve{A}_i = \begin{cases} U_i^{\rm sp} & \text{if } i \in 1 : \bar{m}, \\ V_i^{\rm sp} & \text{if } i \in \bar{m}+1 : m. \end{cases}$$

(If $r < m$, then let $V_{r+1}^{\rm sp}, \ldots, V_m^{\rm sp} \in \mathbb{S}^n$ be such that $\mathrm{range}(\mathcal{A}^*) = \mathrm{span}(\{V_i^{\rm sp} : i \in 1 : m\})$.) Then

$$\mathrm{range}(\breve{\mathcal{A}}^*) \cap \mathrm{range}(Q \cdot Q^\top) = \mathrm{span}(U_1^{\rm sp}, \ldots, U_{\bar{m}}^{\rm sp}) = \mathrm{span}(\breve{A}_1, \ldots, \breve{A}_{\bar{m}}). \tag{5.15}$$

Moreover, define $\breve{\mathcal{V}} : \mathbb{S}^n \to \mathbb{R}^m$ by $\breve{V}_i = V_i^{\rm sp}$, $\forall\, i \in 1 : m$. Then $\mathrm{range}(\mathcal{A}^*) = \mathrm{range}(\breve{\mathcal{V}}^*)$, and $\|(\breve{\mathcal{V}}^* - \breve{\mathcal{A}}^*)y\| \le 2\sqrt{\bar{m}}\varepsilon^{\rm sp}\|y\|$ for all $y \in \mathbb{R}^m$.

Proof. It is immediate that $U_i^{\rm sp} \in \mathrm{range}(\breve{\mathcal{A}}^*) \cap \mathrm{range}(Q \cdot Q^\top)$ for $i \in 1 : \bar{m}$. Since

$$\begin{aligned}
&\max_{\substack{\|U\|=1\\ \|V\|=1}} \left\{ \langle U, V\rangle : U \in \mathrm{range}(Q\cdot Q^\top),\ V \in \mathrm{range}(\breve{\mathcal{A}}^*),\ \langle U, U_j^{\rm sp}\rangle = 0 = \langle V, U_j^{\rm sp}\rangle,\ \forall\, j \in 1 : \bar{m} \right\} \\
\le\ &\max_{\substack{\|U\|=1\\ \|y\|=1}} \left\{ \sum_{i=1}^{\bar{m}} y_i \langle U, U_i^{\rm sp}\rangle + \sum_{i=\bar{m}+1}^{m} y_i \langle U, V_i^{\rm sp}\rangle : U \in \mathrm{range}(Q\cdot Q^\top),\ \langle U, U_j^{\rm sp}\rangle = 0,\ \forall\, j \in 1 : \bar{m} \right\} \\
=\ &\max_{\substack{\|U\|=1\\ \|y\|=1}} \left\{ \sum_{i=\bar{m}+1}^{m} y_i \langle U, V_i^{\rm sp}\rangle : U \in \mathrm{range}(Q\cdot Q^\top),\ \langle U, U_j^{\rm sp}\rangle = 0,\ \forall\, j \in 1 : \bar{m} \right\} \\
=\ &\sigma_{\bar{m}+1}^{\rm sp} < 1 - \varepsilon^{\rm sp} < 1,
\end{aligned}$$

(5.15) holds.

By definition of $V_i^{\rm sp}$'s and $\breve{\mathcal{V}}$, we have $\mathrm{range}(\mathcal{A}^*) = \mathrm{range}(\breve{\mathcal{V}}^*)$. Finally, for any $y \in \mathbb{R}^m$,

$$\|(\breve{\mathcal{V}}^* - \breve{\mathcal{A}}^*)y\| = \left\| \sum_{i=1}^{\bar{m}} (V_i^{\rm sp} - U_i^{\rm sp}) y_i \right\| \le \sum_{i=1}^{\bar{m}} \|V_i^{\rm sp} - U_i^{\rm sp}\| |y_i| \le \sqrt{2\bar{m}\varepsilon^{\rm sp}}\|y\|.$$

□

To increase the robustness of the computation of $\mathrm{range}(\mathcal{A}^*) \cap \mathrm{range}(Q \cdot Q^\top)$, we may follow the treatment in [32], and decide whether a singular value is one by checking its ratios with the previous and the next singular values.

5.4 Shifting the objective

Proposition 5.4.1. *Let* $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m$ *be a one-one linear map satisfying* $\operatorname{range}(\mathcal{A}^*\mathcal{P}) = \operatorname{range}(Q \cdot Q^\top) \cap \operatorname{range}(\mathcal{A}^*)$. *Then the linear equations*

$$\mathcal{A}\mathcal{A}^* y - \mathcal{A}(QQ^\top(\mathcal{A}^* y)QQ^\top) = \mathcal{A}(C - QQ^\top CQQ^\top), \tag{5.16a}$$

$$\mathcal{P}^* y = 0, \tag{5.16b}$$

have a unique solution $\bar{y}$. *Define*

$$\bar{C} := Q^\top(C - \mathcal{A}^*\bar{y})Q \quad \textit{and} \quad C_{\mathrm{res}} := C - \mathcal{A}^*\bar{y} - Q\bar{C}Q^\top;$$

then $(\bar{y}, \bar{C})$ *is an optimal solution of the (underdetermined) linear least squares problem*

$$v_{\mathrm{lss}} = \min_{y,W} \|C - (\mathcal{A}^* y + QWQ^\top)\|. \tag{5.17}$$

Moreover,

$$\mathcal{A}(C_{\mathrm{res}}) = 0, \quad Q^\top C_{\mathrm{res}} Q = 0, \quad \textit{and} \quad v_{\mathrm{lss}} \le \sqrt{2}\left(\min_{Z \in \mathcal{F}_{\mathrm{P}}^Z} \|Z\|\right)\left(\frac{\alpha(\mathcal{A}_C, \delta)\|D\|}{\lambda_{\min}(D_+)}\right). \tag{5.18}$$

Proof. Define the linear function $\mathcal{G} : \mathbb{R}^m \to \mathbb{R}^m$ by

$$\mathcal{G}y := \mathcal{A}\mathcal{A}^* y - \mathcal{A}(QQ^\top(\mathcal{A}^* y)QQ^\top).$$

Then $\mathcal{G}$ is self-adjoint, i.e., $\mathcal{G}^* = \mathcal{G}$. Moreover, $\ker(\mathcal{G}) = \operatorname{range}(\mathcal{P})$.

Now note that $\mathcal{A}(C - QQ^\top CQQ^\top) \in \ker(\mathcal{P}^*) = \operatorname{range}(\mathcal{G}^*) = \operatorname{range}(\mathcal{G})$. Therefore (5.16a) has a solution y. Write $y = y_1 + y_2$, where $y_1 \in \operatorname{range}(\mathcal{G}) = \ker(\mathcal{P}^*)$ and $y_2 \in \ker(\mathcal{G})$; then $\mathcal{L}y = \mathcal{L}y_1$, i.e., y_1 solves (5.16).

To see that (5.16) has a unique solution, simply note that if $y, \tilde{y}$ both solve (5.16), then $y - \tilde{y} \in \ker(\mathcal{G})$ by (5.16a) and $y - \tilde{y} \in \ker(\mathcal{P}^*) = \operatorname{range}(\mathcal{G})$ by (5.16b). Hence $y = \tilde{y}$.

Since the first order optimality condition of (5.17) (which has a convex objective function) is

$$\begin{aligned} \mathcal{A}(C - (\mathcal{A}^* y + QWQ^\top)) &= 0, \\ Q^\top(C - (\mathcal{A}^* y + QWQ^\top))Q &= 0, \end{aligned} \tag{5.19}$$

$(\bar{y}, \bar{C})$ satisfies the optimality condition and hence solves (5.17). That $\mathcal{A}(C_{\mathrm{res}}) = 0$ and $Q^\top C_{\mathrm{res}} Q = 0$ follows immediately from the definition of C_{res} and the optimality condition (5.19).

Finally, let $y \in \mathcal{F}_{\mathrm{P}}^Z$. Then $v_{\mathrm{lss}} \le \|C - \mathcal{A}^* y - QQ^\top(C - \mathcal{A}^* y)QQ^\top\|$, and we can use (5.4) to get (5.18). □

5.5 Numerical results

In this section, we consider using the facial reduction algorithm on semidefinite programs that fail the Slater condition. (Some examples can be found in Section 7.2.)

We quote the result from [27] in Table 5.1 on Page 81, which compares solving 18 instances of the SDP (P) *with* versus *without* using the facial reduction algorithm. Examples 2 to 9 are instances where the true optimal values of (P) and (D) are known. Specifically,

- Example 2 has a positive duality gap: $v_{\mathrm{P}} = 0 < v_{\mathrm{D}} = 1$;
- in Examples 4, 9a and 9b, the dual (D) is infeasible; Examples 9a and 9b uses the problem data defined in (4.17) (so they require $n - 1$ iterations of facial reduction, where n is the size of the matrix variable).

The instances RandGen1-RandGen11 are generated randomly using Algorithm 5.4 on Page 80. Most of them have a finite positive duality gap, because of the following result (with a proof in [27, Theorem 12.39]):

Theorem 5.5.1. *[91, 97] Given any positive integers n, $m \leq n(n+1)/2$ and any $g > 0$ as input for Algorithm 5.4, the following statements hold for the primal-dual pair* (P)-(D) *corresponding to the output data from Algorithm 5.4:*

1. *Both* (P) *and* (D) *are feasible.*
2. *All primal feasible points are optimal and $v_{\mathrm{P}} = 0$.*
3. *All dual feasible point are optimal and $v_{\mathrm{D}} = g > 0$.*

It follows that (P) *and* (D) *possess a finite positive duality gap.*

Moreover, these instances generically require only one iteration of facial reduction. SeDuMi [85] is used to solve the SDPs in both cases.

When the instance has zero duality gap (as in Examples 1, 3, 6 and 7), SeDuMi is able to compute the optimal value. However, when there is a finite nonzero duality gap, SeDuMi may not always be able to solve the SDP, and returns `NaN`. We note that, theoretically, the failure of the Slater condition in a given SDP should not be an issue for the self-dual embedding method. It is not clear why SeDuMi has difficulty handling instances where a nonzero duality gap is present.

Algorithm 5.4: Generating an SDP instance that has a finite nonzero duality gap [91, 97]

1 Input(*problem dimensions m, n; desired duality gap $g > 0$*);

2 Output(*linear map $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, $b \in \mathbb{R}^m$, $C \in \mathbb{S}^n$ such that the corresponding primal dual pair* (P)-(D) *has a finite nonzero duality gap*);

(1) Pick any positive integer r_1,r_3 that satisfy $r_1 + r_3 + 1 = n$,
and any positive integer $p \leq r_3$;

(2) choose $A_i \succeq 0$ for $i = 1, \ldots, p$ so that $\dim(\text{face}(\{A_i : i = 1, \ldots, p\})) = r_3$.
Specifically, choose $A_1, \ldots, A_p$ so that

$$\text{face}(\{A_i : 1, \ldots, p\}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \mathbb{S}_+^{r_3} \end{bmatrix};$$

(3) choose $A_{p+1}, \ldots, A_m$ of the form

$$A_i = \begin{bmatrix} 0 & 0 & (A_i)_{13} \\ 0 & (A_i)_{22} & * \\ ((A_i)_{13})^\top & * & * \end{bmatrix},$$

where an asterisk denotes a block having arbitrary elements, such that

- $(A_{p+1})_{13}, \ldots, (A_m)_{13}$ are linearly independent, and
- $(A_i)_{22} \succ 0$ for some $i \in \{p+1, \ldots, m\}$;

(4) pick

$$\bar{X} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \sqrt{g} & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

(5) take $b = \mathcal{A}(\bar{X})$, $C = \bar{X}$.

Name	n	m	True primal optimal value	True dual optimal value	Primal optimal value with facial reduction	Primal optimal value without facial reduction
Example 1	3	2	0	0	0	-6.30238e-016
Example 2	3	2	0	1	0	+0.570395
Example 3	3	4	0	0	0	+6.91452e-005
Example 4	3	3	0	Infeas.	0	+Inf
Example 6	6	8	1	1	+1	+1
Example 7	5	3	0	0	0	-2.76307e-012
Example 9a	20	20	0	Infeas.	0	Inf
Example 9b	100	100	0	Infeas.	0	Inf
RandGen1	10	5	0	1.4509	+1.5914e-015	+1.16729e-012
RandGen2	100	67	0	5.5288e+003	+1.1056e-010	NaN
RandGen4	200	140	0	2.6168e+004	+1.02803e-009	NaN
RandGen5	120	45	0	0.0381	-5.47393e-015	-1.63758e-015
RandGen6	320	140	0	2.5869e+005	+5.9077e-025	NaN
RandGen7	40	27	0	168.5226	-5.2203e-029	+5.64118e-011
RandGen8	60	40	0	4.1908	-2.03227e-029	NaN
RandGen9	60	40	0	61.0780	+5.61602e-015	-3.52291e-012
RandGen10	180	100	0	5.1461e+004	+2.47204e-010	NaN
RandGen11	255	150	0	4.6639e+004	+7.71685e-010	NaN

Table 5.1: Comparisons with/without facial reduction, using SeDuMi, from [27]

Chapter 6

Backward stability of facial reduction on SDP

In this chapter, we discuss the backward stability of the facial reduction algorithm applied on SDP.

We first recall the notion of backward stability. Suppose that an algorithm takes an input x and calculates $y = f(x)$, where f is some function. While the computed output y from the algorithm may not equal the true value $y_{\text{true}} = f(x)$, it is often considered acceptable as long as $y = f(x')$ for some x' near x. In such a case, we say that the algorithm is *backward stable*. Examples illustrating the idea of backward stability can be found in, e.g., [53].

The facial reduction algorithm on the SDP (P) takes as input the data $(\mathcal{A}, b, C)$ and outputs $(\bar{\mathcal{A}}, \bar{b}, \bar{C})$ such that (P) is equivalent to

$$\sup_v \left\{ \bar{b}^\top v : \bar{C} - \bar{\mathcal{A}}^* v \succeq 0 \right\}, \tag{6.1}$$

which satisfies the Slater condition. (It could be either $(\bar{\mathcal{A}}, \bar{b}, \bar{C}) = (\mathcal{A}, b, C)$ or $(\bar{\mathcal{A}}, \bar{b}, \bar{C})$ is from (4.13), if $(\mathrm{P}_{\text{conic}})$ fails the Slater condition.) As pointed out in the previous chapter, the computation involves a number of nontrivial steps and may incur numerical errors.

In this chapter, we are concerned specifically with the fact that we can only solve the auxiliary problem (5.1) approximately, which may lead to inaccuracy in computing the smaller equivalent problem. In this light, we study the backward stability of one iteration of the facial reduction algorithm in the presence of small error in solving (5.1): suppose that we input $(\mathcal{A}, b, C)$ and that, after one iteration of the facial reduction algorithm (i.e., Algorithm 6.1), we get the computed

ouput $(\bar{\mathcal{A}}, \bar{b}, \bar{C}, \bar{y}, \mathcal{P}, Q)$. We show that this computed output is the true output of Algorithm 6.1 applied on some data $(\tilde{\mathcal{A}}, \tilde{b}, \tilde{C})$ that is not "too far" from $(\mathcal{A}, b, C)$.

We first prove a technical lemma in Section 6.1, that studies the minimum singular value of a one-one linear map. (Section 6.1 may be skipped on the first reading.) In Section 6.2, we state and prove the backward stability result for one iteration of the facial reduction.

6.1 A technical lemma

Lemma 6.1.1. *Following the notation and assumptions of Theorem 5.3.1, and extending the set* $\{V_1^{\rm sp}, \ldots, V_r^{\rm sp}\}$ *to an orthonormal basis* $\{V_1^{\rm sp}, \ldots, V_m^{\rm sp}\}$ *of* $\mathrm{range}(\mathcal{A}^*)$ *if* $r < m$*, define linear maps* $\mathcal{V}, \mathcal{H} : \mathbb{S}^n \to \mathbb{R}^{m-\bar{m}}$ *by*

$$\mathcal{V}^* v := \sum_{i=1}^{m-\bar{m}} v_i V_{\bar{m}+i}^{\rm sp} \quad \textit{and} \quad \mathcal{H}^* v := \mathcal{V}^* v - QQ^\top(\mathcal{V}^* v)QQ^\top, \quad \forall\, v \in \mathbb{R}^{m-\bar{m}}.$$

Then

$$\sigma_{\min}(\mathcal{H}^*) = \sqrt{1 - (\sigma_{\bar{m}+1}^{\rm sp})^2} > 0, \tag{6.2}$$

where $\sigma_i^{\rm sp}$ *is defined in* (5.12) *for all* $i \in 1 : \min\left\{\frac{1}{2}\bar{n}(\bar{n}+1), m\right\}$.

Proof. We first prove that

$$\sigma_{\bar{m}+1}^{\rm sp} = \|Q^\top V_{\bar{m}+1}^{\rm sp} Q\| \quad \text{and} \quad U_{\bar{m}+1}^{\rm sp} = QQ^\top V_{\bar{m}+1}^{\rm sp} QQ^\top, \tag{6.3}$$

where $U_i^{\rm sp} \in \mathrm{range}(Q \cdot Q^\top)$ defines the principal angle $\sigma_i^{\rm sp}$ for $i \in 1 : \min\left\{\frac{1}{2}\bar{n}(\bar{n}+1), m\right\}$; see (5.12) in Theorem 5.3.1.

In fact, by definition,

$$\begin{aligned}
\sigma_{\bar{m}+1}^{\rm sp} = \max_{U,V} \Big\{ &\langle U, V\rangle : U \in \mathrm{range}(Q \cdot Q^\top),\ V \in \mathrm{range}(\mathcal{A}^*), \\
&\qquad \|U\| = 1 = \|V\|,\ \langle U, U_i^{\rm sp}\rangle = 0 = \langle V, V_i^{\rm sp}\rangle,\ \forall\, i \in 1 : \bar{m}\} \\
\geq \max_{\|U\|=1} \Big\{ &\langle U, V_{\bar{m}+1}^{\rm sp}\rangle : U \in \mathrm{range}(Q \cdot Q^\top),\ \langle U, U_i^{\rm sp}\rangle = 0,\ \forall\, i \in 1 : \bar{m} \Big\}.
\end{aligned} \tag{6.4}$$

Since $U_i^{\rm sp} = V_i^{\rm sp}$ for $i \in 1 : \bar{m}$, we have

$$\langle QQ^\top V_{\bar{m}+1}^{\rm sp} QQ^\top, U_i^{\rm sp}\rangle = \langle V_{\bar{m}+1}^{\rm sp}, U_i^{\rm sp}\rangle = \langle V_{\bar{m}+1}^{\rm sp}, V_i^{\rm sp}\rangle = 0,\ \forall\, i \in 1 : \bar{m},$$

i.e., $U = \frac{1}{\|Q^\top V_{\bar{m}+1}^{\rm sp} Q\|} QQ^\top V_{\bar{m}+1}^{\rm sp} QQ^\top \in \mathrm{range}(Q \cdot Q^\top)$ is feasible for (6.4). Therefore

$$\sigma_{\bar{m}+1}^{\rm sp} \geq \|Q^\top V_{\bar{m}+1}^{\rm sp} Q\| = \|QQ^\top V_{\bar{m}+1}^{\rm sp} QQ^\top\| \|U_{\bar{m}+1}^{\rm sp}\| \geq \langle QQ^\top V_{\bar{m}+1}^{\rm sp} QQ^\top, U_{\bar{m}+1}^{\rm sp}\rangle = \sigma_{\bar{m}+1}^{\rm sp},$$

implying (6.3).

Now we prove (6.2). Since for all $v \in \mathbb{R}^{m-\bar{m}}$,

$$\|\mathcal{H}^* v\|^2 = \|\mathcal{V}^* v\|^2 - \|Q^\top(\mathcal{V}^* v)Q\|^2, \quad \text{and} \quad \|\mathcal{V}^* v\| = \|v\|$$

(because $V^{\rm sp}_{\bar{m}+1}, \ldots, V^{\rm sp}_m$ are orthonormal to each other), we have

$$\begin{aligned} \sigma_{\min}(\mathcal{H}^*)^2 = & \min_v \Big\{ \|\mathcal{V}^* v\|^2 - \|Q^\top(\mathcal{V}^* v)Q\|^2 : \|v\| = 1 \Big\} \\ = & \min_v \Big\{ \|\mathcal{V}^* v\|^2 - \|Q^\top(\mathcal{V}^* v)Q\|^2 : \|\mathcal{V}^* v\| = 1 \Big\} \\ = & \ 1 - \max_v \Big\{ \|Q^\top(\mathcal{V}^* v)Q\|^2 : \|\mathcal{V}^* v\| = 1 \Big\}. \end{aligned} \tag{6.5}$$

Observe that

$$V \in \text{range}(\mathcal{V}^*) \iff V \in \text{range}(\mathcal{A}^*),\ \langle V, V_i^{\rm sp}\rangle = 0,\ \forall\, i \in 1 : \bar{m};$$

therefore

$$\begin{aligned} & \max_v \Big\{ \|Q^\top(\mathcal{V}^* v)Q\| : \|\mathcal{V}^* v\| = 1 \Big\} \\ = & \max_V \Big\{ \|Q^\top V Q\| : V \in \text{range}(\mathcal{A}^*),\ \|V\| = 1,\ \langle V, V_i^{\rm sp}\rangle = 0,\ \forall\, i \in 1 : \bar{m} \Big\} \\ \geq & \ \|Q^\top V^{\rm sp}_{\bar{m}+1} Q\|. \end{aligned} \tag{6.6}$$

On the other hand,

$$\begin{aligned} & \max_v \Big\{ \|Q^\top(\mathcal{V}^* v)Q\| : \|\mathcal{V}^* v\| = 1 \Big\} \\ = & \max_{v,U} \Big\{ \langle U, \mathcal{V}^* v\rangle : \|\mathcal{V}^* v\| = 1,\ U \in \text{range}(Q \cdot Q^\top),\ \|U\| = 1 \Big\} \qquad (6.7) \\ \geq & \max_{U,V} \Big\{ \langle U, V\rangle : U \in \text{range}(Q \cdot Q^\top),\ V \in \text{range}(\mathcal{V}^*), \qquad (6.8) \\ & \qquad\qquad \|U\| = 1 = \|V\|,\ \langle U, U_i^{\rm sp}\rangle = 0,\ \forall\, i = 1 : \bar{m} \Big\} \end{aligned}$$

Let (v^*, U^*) be an optimal solution of (6.7). Write $U^* = \sum_{i=1}^{\bar{m}} u_i U_i^{\rm sp} + \tilde{U}$, where $\langle \tilde{U}, U_i^{\rm sp}\rangle = 0$ for all $i \in 1 : \bar{m}$. We show that $u_i = 0$ for all $i \in 1 : \bar{m}$. Note that $\langle \mathcal{V}^* v^*, U_i^{\rm sp}\rangle = \langle \mathcal{V}^* v^*, V_i^{\rm sp}\rangle = 0$ for all $i \in 1 : \bar{m}$, i.e., $\langle U, \mathcal{V}^*\rangle = \langle \tilde{U}, \mathcal{V}^*\rangle$. If $u_i \neq 0$ for some $i \in 1 : \bar{m}$, then $\|\tilde{U}\| < \|U\| = 1$. Then $(\tilde{U}/\|\tilde{U}\|, \mathcal{V}^* v^*)$ is a feasible solution of (6.8) and its objective value is $\langle U, \mathcal{V}^*\rangle/\|\tilde{U}\| > \langle U, \mathcal{V}^*\rangle$, which contradicts the inequality in (6.8). Therefore we must have that $U^* = \tilde{U}$, proving that the optimization problems (6.7) and (6.8) have the same optimal value. Therefore, combining with

(6.6),

$$\begin{aligned}\|Q^\top V^{\rm sp}_{\bar m+1}Q\| \le & \max_{U,V}\Big\{\langle U,V\rangle : U\in \mathrm{range}(Q\cdot Q^\top),\ V\in\mathrm{range}(\mathcal{V}^*),\\ & \qquad\qquad \|U\|=1=\|V\|,\ \langle U,U_i^{\rm sp}\rangle=0,\ \forall\, i=1:\bar m\Big\}\\ = & \max_{U,V}\Big\{\langle U,V\rangle : U\in \mathrm{range}(Q\cdot Q^\top),\ V\in\mathrm{range}(\mathcal{A}^*),\\ & \qquad\qquad \|U\|=1=\|V\|,\ \langle U,U_i^{\rm sp}\rangle=0=\langle V,V_i^{\rm sp}\rangle,\ \forall\, i=1:\bar m\Big\}\\ = & \sigma^{\rm sp}_{\bar m+1}=\|Q^\top V^{\rm sp}_{\bar m+1}Q\|,\end{aligned}$$

where the last equality follows from (6.3). Therefore we have

$$\sigma^{\rm sp}_{\bar m+1}=\max_v\Big\{\|Q^\top(\mathcal{V}^*v)Q\|^2 : \|\mathcal{V}^*v\|=1\Big\},$$

which together with (6.5) implies (6.2). □

6.2 Backward stability of one iteration of facial reduction

In this section, we show that one iteration of the facial reduction (i.e., Algorithm 6.1 on Page 87) is backward stable. We first state the backward stability result:

Theorem 6.2.1. *[27, Theorem 12.38] Let $b\in\mathbb{R}^m$, $C\in\mathbb{S}^n$ and an onto linear map $\mathcal{A}:\mathbb{S}^n\to\mathbb{R}^m$ be given so that* (P) *is feasible. Suppose that Algorithm 6.1 finds a feasible solution (δ, D) of the auxiliary problem* (5.1) *in Step (1) with*

$$\delta\ge 0,\quad D=\begin{bmatrix}P & Q\end{bmatrix}\begin{bmatrix}D_+ & 0\\ 0 & D_\epsilon\end{bmatrix}\begin{bmatrix}P^\top\\ Q^\top\end{bmatrix}\succeq 0$$
$$\text{with}\ \ D_+\in\mathbb{S}^{n-\bar n}_{++}\ (0<\bar n<n),\quad \begin{bmatrix}P & Q\end{bmatrix}\in\mathbb{R}^{n\times n}\ \ \text{orthogonal},$$

and returns $(\bar{\mathcal{A}},\bar b,\bar C,\bar y,\mathcal{P})$. In addition, assume that

$$(\sigma^{\rm sp}_{\bar m+1})^2<1-\frac{2(m-\bar m)}{\|D_+\|^2}\left(\frac{\|\mathcal{A}(D)\|^2}{\sigma_{\min}(\mathcal{A}^*)^2}+\|D_\epsilon\|^2\right)$$

holds. Then $(\bar{\mathcal{A}},\bar b,\bar C)$is the exact output of Algorithm 6.1 applied on $(\tilde{\mathcal{A}},b,\tilde C)$, where $\tilde{\mathcal{A}}:\mathbb{S}^n\to\mathbb{R}^m:(\langle\tilde A_i,X\rangle)$ is defined by

$$\tilde A_i:=A_i-\frac{\langle A_i,PD_+P^\top\rangle}{\|D_+\|^2}PD_+P^\top,$$

and $\tilde C:=\tilde{\mathcal{A}}^*\bar y+Q\bar CQ^\top$. In other words, the following hold:

(1) $\tilde{\mathcal{A}}(PD_+P^\top) = 0$, $\langle \tilde{C}, PD_+P^\top\rangle = 0$.

(2) $\operatorname{range}(\tilde{\mathcal{A}}^*\mathcal{P}) = \operatorname{range}(Q\cdot Q^\top) \cap \operatorname{range}(\tilde{\mathcal{A}}^*)$.

(3) $\bar{y}$ *solves*

$$\begin{aligned} \tilde{\mathcal{A}}\tilde{\mathcal{A}}^*\bar{y} - \tilde{\mathcal{A}}(QQ^\top(\tilde{\mathcal{A}}^*y)QQ^\top) = &\quad \tilde{\mathcal{A}}(\tilde{C} - QQ^\top\tilde{C}QQ^\top), \\ \mathcal{P}^*\bar{y} = &\quad 0, \end{aligned} \tag{6.9}$$

and $(\bar{y}, \bar{C})$ *solves the least squares problem*

$$\min_{y,W} \|\tilde{\mathcal{A}}^*y + QWQ^\top - \tilde{C}\|. \tag{6.10}$$

Moreover,

$$\|\mathcal{A}^* - \tilde{\mathcal{A}}^*\| \leq \|\mathcal{A}(D)\| + \|D_\epsilon\| \left(\sum_{i=1}^m \|A_i\|^2\right)^{1/2},$$

$$\text{and} \qquad \|C - \tilde{C}\| \leq \sqrt{2}\left(\min_{Z\in\mathcal{F}_{\mathrm{P}}^Z} \|Z\|\right)\left(\frac{\alpha(\mathcal{A}_C,\delta)\|D\|}{\lambda_{\min}(D_+)}\right).$$

Before we prove Theorem 6.2.1, we need a lemma about finding a linear map $\tilde{\mathcal{A}}$ that is "near" $\mathcal{A}$ that satisfies

$$\operatorname{range}(Q\cdot Q^\top) \cap \operatorname{range}(\tilde{\mathcal{A}}^*) = \operatorname{range}(Q\cdot Q^\top) \cap \operatorname{range}(\mathcal{A}^*).$$

Lemma 6.2.2. *Let* $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m : X \mapsto (\langle A_i, X\rangle)$ *be linear onto.*

Let $D = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & D_\epsilon \end{bmatrix} \begin{bmatrix} P^\top \\ Q^\top \end{bmatrix} \in \mathbb{S}^n_+$, *where* $\begin{bmatrix} P & Q \end{bmatrix} \in \mathbb{R}^{n\times n}$ *is orthogonal and* $D_+ \succ 0$, *and let* $\sigma_1^{\mathrm{sp}}, \sigma_2^{\mathrm{sp}}, \ldots, \sigma^{\mathrm{sp}}_{\min\left\{\frac{1}{2}\bar{n}(\bar{n}+1),m\right\}} \geq 0$ *satisfy* (5.12), *with* $\sigma^{\mathrm{sp}}_{\bar{m}} = 1 > \sigma^{\mathrm{sp}}_{\bar{m}+1}$. *Assume that*

$$1 - (\sigma^{\mathrm{sp}}_{\bar{m}+1})^2 > \frac{2(m-\bar{m})}{\|D_+\|^2}\left(\frac{\|\mathcal{A}(D)\|^2}{\sigma_{\min}(\mathcal{A}^*)^2} + \|D_\epsilon\|^2\right). \tag{6.11}$$

Define $\tilde{A}_i$ *to be the projection of* A_i *onto* $\left\{PD_+P^\top\right\}^\perp$:

$$\tilde{A}_i := A_i - \frac{\langle A_i, PD_+P^\top\rangle}{\|D_+\|^2} PD_+P^\top, \ \forall\, i \in 1:m. \tag{6.12}$$

and $\tilde{\mathcal{A}} : \mathbb{S}^n \to \mathbb{R}^m$ *by* $\tilde{\mathcal{A}}^*y := \sum_{i=1}^m y_i\tilde{A}_i$. *Then*

$$\operatorname{range}(Q\cdot Q^\top) \cap \operatorname{range}(\tilde{\mathcal{A}}^*) = \operatorname{range}(Q\cdot Q^\top) \cap \operatorname{range}(\mathcal{A}^*). \tag{6.13}$$

Algorithm 6.1: One iteration of the facial reduction algorithm

Input($\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$, $b \in \mathbb{R}^m$, $C \in \mathbb{S}^n$ *such that* (P) *is feasible.*)

Step (1): solve the auxiliary problem

perform preprocessing on (5.1);

obtain an optimal solution of (δ^*, D^*) of (5.1);

Step (2): find a smaller face of $\mathbb{S}^n_+$ containing $\mathcal{F}_{\rm P}^Z$

if $D^* = 0$ *or* $D^* \succ 0$ **then**

    STOP;

else

    obtain spectral decomposition

$$D^* = \begin{bmatrix} P & Q \end{bmatrix} \begin{bmatrix} D_+ & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} P^\top \\ Q^\top \end{bmatrix}, \quad \text{with } D_+ \succ 0 \text{ and } Q \in \mathbb{R}^{n \times \bar{n}};$$

endif

Step (3): compute the subspace intersection

find $\sigma_1^{\rm sp}, \ldots, \sigma_r^{\rm sp} \geq 0$ and $V_1^{\rm sp}, \ldots, V_r^{\rm sp} \in \text{range}(\mathcal{A}^*)$ satisfying (5.12) via Algorithm 5.3;

if $\sigma_1^{\rm sp} < 1$ **then**

    STOP; all feasible solutions are optimal;

else

    let $\bar{m}$ satisfy $\sigma_1^{\rm sp} \geq \sigma_1^{\rm sp} \geq \cdots \geq \sigma_{\bar{m}}^{\rm sp} = 1 > \sigma_{\bar{m}+1}^{\rm sp} \geq 0$;

    define the linear map $\mathcal{P}v := \sum_{i=1}^{\bar{m}} v_i (\mathcal{A}^*)^\dagger (V_i^{\rm sp})$ for all $v \in \mathbb{R}^{\bar{m}}$;

endif

Step (4): shifting the objective

solve (5.16) for $\bar{y}$;

Step (5): project the problem data

$\bar{\mathcal{A}}^*(\cdot) \leftarrow Q^\top (\mathcal{A}^* \mathcal{P}(\cdot)) Q$;

$\bar{b} \leftarrow \mathcal{P}^* b$;

$\bar{C} \leftarrow Q^\top (C - \mathcal{A}^* \bar{y}) Q$.

Output($\bar{\mathcal{A}}, \bar{b}, \bar{C}$)

Proof. Define $r := \min\left\{\frac{1}{2}\bar{n}(\bar{n}+1), m\right\}$; let $U_1^{\mathrm{sp}}, U_2^{\mathrm{sp}}, \ldots, U_r^{\mathrm{sp}} \in \mathrm{range}(Q \cdot Q^\top)$, $V_1^{\mathrm{sp}}, V_2^{\mathrm{sp}}, \ldots, V_r^{\mathrm{sp}} \in \mathrm{range}(\mathcal{A}^*)$ along with $\sigma_1^{\mathrm{sp}}, \sigma_2^{\mathrm{sp}}, \ldots, \sigma_r^{\mathrm{sp}} \geq 0$ satisfy (5.12). Then

$$\mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\mathcal{A}^*) = \mathrm{range}(\mathcal{A}^*\mathcal{P}),$$

where $\mathcal{P} : \mathbb{R}^{\bar{m}} \to \mathbb{R}^m : v \mapsto \sum_{i=1}^{\bar{m}} v_i (\mathcal{A}^*)^\dagger (V_i^{\mathrm{sp}})$.

First note that by definition of $\tilde{A}_i$'s in (6.12), for any $y \in \mathbb{R}^m$,

$$\tilde{\mathcal{A}}^* y = \mathcal{A}^* y - \frac{\langle \mathcal{A}^* y, P D_+ P^\top \rangle}{\|D_+\|^2} P D_+ P^\top. \tag{6.14}$$

It is easy to see that $\mathrm{range}(\mathcal{A}^*\mathcal{P}) \subseteq \mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\tilde{\mathcal{A}}^*)$: for any $u \in \mathbb{R}^{\bar{m}}$, $\langle \mathcal{A}^*\mathcal{P}u, P D_+ P^\top \rangle = 0$, hence $\mathcal{A}^*\mathcal{P}u = \tilde{\mathcal{A}}^*\mathcal{P}u \in \mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\tilde{\mathcal{A}}^*)$.

For the converse, fix any $y \in \mathbb{R}^m$ such that $\tilde{\mathcal{A}}^* y \in \mathrm{range}(Q \cdot Q^\top)$; we show that $\tilde{\mathcal{A}}^* y = \mathcal{A}^* y \in \mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\mathcal{A}^*)$. Extend $\{V_1^{\mathrm{sp}}, \ldots, V_r^{\mathrm{sp}}\}$ to an orthonormal basis $\{V_1^{\mathrm{sp}}, \ldots, V_m^{\mathrm{sp}}\}$ if $r < m$; then $\mathcal{A}^* y = \sum_{i=1}^m v_i V_i^{\mathrm{sp}}$ for some $v \in \mathbb{R}^m$. We prove that $v_i = 0$ for all $i \in (\bar{m}+1) : m$.

The inclusion $\tilde{\mathcal{A}}^* y \in \mathrm{range}(Q \cdot Q^\top)$ implies that $QQ^\top(\tilde{\mathcal{A}}^* y)QQ^\top = \tilde{\mathcal{A}}^* y$, and by (6.14) we get

$$QQ^\top(\mathcal{A}^* y)QQ^\top = \mathcal{A}^* y - \frac{\langle \mathcal{A}^* y, P D_+ P^\top \rangle}{\|D_+\|^2} P D_+ P^\top. \tag{6.15}$$

Since $\mathcal{A}^* y = \sum_{i=1}^m v_i V_i^{\mathrm{sp}}$ and $\sum_{i=1}^{\bar{m}} v_i V_i^{\mathrm{sp}} \in \mathrm{range}(Q \cdot Q^\top)$, (6.15) implies that

$$\sum_{i=\bar{m}+1}^{m} v_i \left(V_i^{\mathrm{sp}} - QQ^\top V_i^{\mathrm{sp}} QQ^\top \right) = \frac{\langle \sum_{i=\bar{m}+1}^m v_i V_i^{\mathrm{sp}}, P D_+ P^\top \rangle}{\|D_+\|^2} P D_+ P^\top. \tag{6.16}$$

Taking inner product with V_j^{sp} on both sides of (6.16) for $j \in (\bar{m}+1) : m$:

$$\sum_{i=\bar{m}+1}^{m} v_i \left(\langle V_i^{\mathrm{sp}}, V_j^{\mathrm{sp}} \rangle - \langle Q^\top V_i^{\mathrm{sp}} Q,\, Q^\top V_j^{\mathrm{sp}} Q \rangle \right) = \sum_{i=\bar{m}+1}^{m} v_i \frac{\langle V_i^{\mathrm{sp}}, P D_+ P^\top \rangle \langle V_j^{\mathrm{sp}}, P D_+ P^\top \rangle}{\|D_+\|^2},$$

which holds for all $j \in (\bar{m}+1) : m$ if and only if

$$(M - \tilde{M}) \begin{pmatrix} v_{\bar{m}+1} \\ \vdots \\ v_m \end{pmatrix} = 0, \tag{6.17}$$

where $M, \tilde{M} \in \mathbb{S}^{m-\bar{m}}$ are defined by

$$\begin{aligned} M_{(i-\bar{m}),(j-\bar{m})} &= \langle V_i^{\mathrm{sp}}, V_j^{\mathrm{sp}} \rangle - \langle Q^\top V_i^{\mathrm{sp}} Q,\, Q^\top V_j^{\mathrm{sp}} Q \rangle, \\ \tilde{M}_{(i-\bar{m}),(j-\bar{m})} &= \frac{\langle V_i^{\mathrm{sp}}, P D_+ P^\top \rangle \langle V_j^{\mathrm{sp}}, P D_+ P^\top \rangle}{\|D_+\|^2}, \qquad \forall\, i, j \in (\bar{m}+1) : m. \end{aligned}$$

We show that $(v_{\bar{m}+1}, \ldots, v_m)$ solves (6.17) if and only if $v_{\bar{m}+1} = \cdots = v_m$ by proving that $\lambda_{\min}(M - \tilde{M}) \geq \lambda_{\min}(M) - \lambda_{\max}(\tilde{M}) > 0$ holds if we assume that (6.11) holds.

First we estimate $\lambda_{\min}(M)$:

$$\begin{aligned}
\lambda_{\min}(M) &= \min_{\|u\|=1} \left\{ v^\top M v \right\} = \min_{\|u\|=1} \left\{ \sum_{i,j=1}^{m-\bar{m}} M_{ij} u_i u_j \right\} \\
&= \min_{\|u\|=1} \left\{ \left\langle \sum_{i=1}^{m-\bar{m}} u_i V^{\mathrm{sp}}_{\bar{m}+i} \,,\, \sum_{j=1}^{m-\bar{m}} u_j V^{\mathrm{sp}}_{\bar{m}+j} \right\rangle - \left\langle \sum_{i=1}^{m-\bar{m}} u_i Q^\top V^{\mathrm{sp}}_{\bar{m}+i} Q \,,\, \sum_{j=1}^{m-\bar{m}} u_j Q^\top V^{\mathrm{sp}}_{\bar{m}+j} Q \right\rangle \right\} \\
&= \min_{\|u\|=1} \left\{ \left\| \sum_{i=1}^{m-\bar{m}} u_i V^{\mathrm{sp}}_{\bar{m}+i} \right\|^2 - \left\| Q^\top \left(\sum_{i=1}^{m-\bar{m}} u_i V^{\mathrm{sp}}_{\bar{m}+i} \right) Q \right\|^2 \right\}; \\
&= \min_{\|u\|=1} \left\{ \left\| \sum_{i=1}^{m-\bar{m}} u_i \left(V^{\mathrm{sp}}_{\bar{m}+i} - Q Q^\top V^{\mathrm{sp}}_{\bar{m}+i} Q Q^\top \right) \right\|^2 \right\} \\
&= 1 - (\sigma^{\mathrm{sp}}_{\bar{m}+1})^2 \,,
\end{aligned}$$

where the last equality follows from Lemma 6.1.1. Next we estimate $\lambda_{\max}(\tilde{M})$:

$$\begin{aligned}
\lambda_{\max}(\tilde{M}) &= \max_{\|u\|=1} \left\{ u^\top \tilde{M} u \right\} = \max_{\|u\|=1} \left\{ \frac{1}{\|D_+\|^2} \sum_{i,j=1}^{m-\bar{m}} u_i u_j \langle V^{\mathrm{sp}}_{\bar{m}+i}, P D_+ P^\top \rangle \langle V^{\mathrm{sp}}_{\bar{m}+j}, P D_+ P^\top \rangle \right\} \\
&= \frac{1}{\|D_+\|^2} \max_{\|u\|=1} \left\{ \left(\sum_{j=1}^{m-\bar{m}} u_j \langle V^{\mathrm{sp}}_{\bar{m}+j}, P D_+ P^\top \rangle \right)^2 \right\} \\
&\leq \frac{1}{\|D_+\|^2} \max_{\|u\|=1} \left\{ \|u\|^2 \sum_{j=1}^{m-\bar{m}} \langle V^{\mathrm{sp}}_{\bar{m}+j}, P D_+ P^\top \rangle^2 \right\} = \frac{1}{\|D_+\|^2} \sum_{j=1}^{m-\bar{m}} \langle V^{\mathrm{sp}}_{\bar{m}+j}, P D_+ P^\top \rangle^2.
\end{aligned}$$

Now note that since $D = P D_+ P^\top + Q D_\epsilon Q^\top$, for each $j \in (\bar{m}+1) : m$,

$$\begin{aligned}
\left| \langle V^{\mathrm{sp}}_{\bar{m}+j} \,,\, P D_+ P^\top \rangle \right| &\leq \left| \langle V^{\mathrm{sp}}_{\bar{m}+j} \,,\, D \rangle \right| + \left| \langle V^{\mathrm{sp}}_{\bar{m}+j} \,,\, Q D_\epsilon Q^\top \rangle \right| \\
&\leq \left| \langle V^{\mathrm{sp}}_{\bar{m}+j} \,,\, D \rangle \right| + \|V^{\mathrm{sp}}_{\bar{m}+j}\| \|Q D_\epsilon Q^\top\| \\
&= \left| \langle V^{\mathrm{sp}}_{\bar{m}+j} \,,\, D \rangle \right| + \|D_\epsilon\| \\
&\leq \sqrt{2} \left(|\langle V^{\mathrm{sp}}_{\bar{m}+j} \,,\, D \rangle|^2 + \|D_\epsilon\|^2 \right)^{1/2}.
\end{aligned}$$

For each $j \in (\bar{m}+1) : m$, $V_j^{\rm sp} = \mathcal{A}^* y^{(j)}$ and $\|y^{(j)}\| \le \frac{1}{\sigma_{\min}(\mathcal{A}^*)}$. Hence

$$\begin{aligned}
\lambda_{\max}(\tilde{M}) &\le \frac{2}{\|D_+\|^2} \sum_{j=1}^{m-\bar{m}} \left(|\langle V_{\bar{m}+j}^{\rm sp} \,,\, D\rangle|^2 + \|D_\epsilon\|^2 \right) \\
&= \frac{2}{\|D_+\|^2} \sum_{j=1}^{m-\bar{m}} \left(\left| (y^{(j)})^\top (\mathcal{A}(D)) \right|^2 + \|D_\epsilon\|^2 \right) \\
&\le \frac{2}{\|D_+\|^2} \sum_{j=1}^{m-\bar{m}} \left(\|y^{(j)}\|^2 \|\mathcal{A}(D)\|^2 + \|D_\epsilon\|^2 \right) \\
&\le \frac{2(m-\bar{m})}{\|D_+\|^2} \left(\frac{\|\mathcal{A}(D)\|^2}{\sigma_{\min}(\mathcal{A}^*)^2} + \|D_\epsilon\|^2 \right).
\end{aligned}$$

Therefore we have

$$\begin{aligned}
\lambda_{\min}(M - \tilde{M}) &\ge \lambda_{\min}(M) - \lambda_{\max}(\tilde{M}) \\
&\ge 1 - (\sigma_{\bar{m}+1}^{\rm sp})^2 - \frac{2(m-\bar{m})}{\|D_+\|^2} \left(\frac{\|\mathcal{A}(D)\|^2}{\sigma_{\min}(\mathcal{A}^*)^2} + \|D_\epsilon\|^2 \right) \\
&> 0,
\end{aligned}$$

where the last inequality follows from the assumption (6.11). Therefore $M - \tilde{M}$ is positive definite, and (6.17) implies that $v_{\bar{m}+1} = v_{\bar{m}+2} = \cdots = v_m = 0$. Therefore $\mathcal{A}^* y = \sum_{i=1}^m v_i V_i^{\rm sp} = \sum_{i=1}^{\bar{m}} v_i V_i^{\rm sp} \in \mathrm{range}(Q \cdot Q^\top)$. Therefore $\tilde{\mathcal{A}}^* y = \mathcal{A}^* y \in \mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\mathcal{A}^*)$ by (6.14). This shows that $\mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\tilde{\mathcal{A}}^*) \subseteq \mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\mathcal{A}^*)$. Consequently, (6.13) holds. □

Remark. We remark that (6.11) is a mild assumption, so long as the computed solution (δ, D) of the auxiliary problem satisfies $\delta \approx 0$ and D is partitioned properly.

Fix $\varepsilon \in (0,1)$. Suppose that we obtain a computed optimal solution (δ, D) of the auxiliary problem in Step (1) of Algorithm 6.1, and that

$$\delta \le \frac{\sigma_{\min}(\mathcal{A}^*)\varepsilon}{\sqrt{2m}} \quad \left(\Longrightarrow \ \frac{\sqrt{2\,m}\,\delta}{\sigma_{\min}(\mathcal{A}^*)} \le \varepsilon \right);$$

then partitioning D using with the numerical rank decided by the parameter $\gamma := \frac{\varepsilon \|D\|}{\sqrt{n}}$ (see Section 5.1), we have

$$\|D_+\|^2 = \|D\|^2 - \|D_\epsilon\|^2 \ge (1-\varepsilon^2)\|D\|^2 \ge \frac{1-\epsilon^2}{n},$$

and we get

$$\frac{2m}{\|D_+\|^2} \left(\frac{\|\mathcal{A}(D)\|^2}{\sigma_{\min}(\mathcal{A}^*)^2} + \|D_\epsilon\|^2 \right) \le \frac{n\varepsilon^2}{1-\varepsilon^2} + \frac{\varepsilon^2}{1-\varepsilon^2} \le \frac{2n\varepsilon^2}{1-\varepsilon^2}.$$

In particular, if $\bar{m}$ is chosen such that

$$\sigma^{\mathrm{sp}}_{\bar{m}+1} < 1 - \frac{\sqrt{2n}\varepsilon}{\sqrt{1-\varepsilon^2}} \implies (\sigma^{\mathrm{sp}}_{\bar{m}+1})^2 < 1 - \frac{2n\varepsilon^2}{1-\varepsilon^2},$$

then (6.11) holds.

Now we can prove Theorem 6.2.1.

Proof of Theorem 6.2.1. (1): Since each $\tilde{A}_i$ is the projection of A_i onto $\left\{PD_+P^\top\right\}^\top$, we have that $\langle \tilde{A}_i, PD_+P^\top\rangle = 0$. In particular, $\langle \tilde{C}, PD_+P^\top\rangle = \langle Q\bar{C}Q^\top, PD_+P^\top\rangle = 0$.

(2): Let $y^{(i)} := (\mathcal{A}^*)^\dagger V_i^{\mathrm{sp}}$ (so $\mathcal{A}^* y^{(i)} = V_i^{\mathrm{sp}}$). Then for all $i \in 1:\bar{m}$, $\tilde{\mathcal{A}}^* y^{(i)} = \mathcal{A}^* y^{(i)}$ by (6.14). Therefore $\tilde{\mathcal{A}}^*\mathcal{P} = \mathcal{A}^*\mathcal{P}$, and in particular $\mathrm{range}(\tilde{\mathcal{A}}^*\mathcal{P}) = \mathrm{range}(Q \cdot Q^\top) \cap \mathrm{range}(\tilde{\mathcal{A}}^*)$.

(3): Since $\tilde{C} = \tilde{\mathcal{A}}^*\bar{y} + Q\bar{C}Q^\top$, we have $QQ^\top\tilde{C}QQ^\top = QQ^\top\tilde{\mathcal{A}}^*\bar{y}QQ^\top + Q\bar{C}Q^\top$. Therefore

$$\tilde{C} - QQ^\top\tilde{C}QQ^\top = \tilde{\mathcal{A}}^*\bar{y} - QQ^\top\tilde{\mathcal{A}}^*\bar{y}QQ^\top.$$

Recall that $\bar{y}$ satisfies (5.16), so $\mathcal{P}^*\bar{y} = 0$. Therefore $\bar{y}$ solves (6.9). By definition of $\tilde{C}$, we have $\bar{C} = Q^\top(\tilde{C} - \tilde{\mathcal{A}}^*\bar{y})Q$, so by Proposition 5.4.1, $(\bar{y}, \bar{C})$ solves the least squares problem (6.10).

Finally, by (6.14),

$$\begin{aligned}
\|\tilde{\mathcal{A}}^* y - \mathcal{A}^* y\| = |\langle \mathcal{A}^* y,\, PD_+P^\top\rangle| &\le \|y\|\|\mathcal{A}(D - QD_\epsilon Q^\top)\| \\
&\le \|y\|\left(\|\mathcal{A}(D)\| + \left(\sum_{i=1}^m \langle Q^\top A_i Q, D_\epsilon\rangle^2\right)^{1/2}\right) \\
&\le \|y\|\left(\|\mathcal{A}(D)\| + \|D_\epsilon\|\left(\sum_{i=1}^m \|A_i\|^2\right)^{1/2}\right),
\end{aligned}$$

and $\|C - \tilde{C}\| = \|C_{\mathrm{res}}\| \le \sqrt{2}\left(\min_{Z\in\mathcal{F}_{\mathrm{P}}^Z}\|Z\|\right)\left(\frac{\alpha(\mathcal{A}_C,\delta)\|D\|}{\lambda_{\min}(D_+)}\right)$ from Proposition 5.4.1. □

Part III

Applications of the facial reduction

Chapter 7

Sensitivity analysis of SDPs

One interesting theoretical use of facial reduction was discovered by Sturm [86]: given a feasible linear matrix inequality (LMI), its forward and backward errors are related by the number of facial reduction iterations required to find the minimal face of the LMI in question. (This number is called the *degree of singularity* of the LMI in [86]; see Theorem 7.5.1.) In this chapter, we make the following assumption:

Assumption 7.1. *The SDP* (P) *is feasible and has finite optimal value.*

We can show that if (P) is feasible with finite optimal value, then whenever the perturbed problem

$$\mathrm{val_P}(S) := \sup_y \left\{ b^\top y : C - \mathcal{A}^* y \succeq S \right\}, \tag{7.1}$$

is feasible and S is small, we have

$$\mathrm{val_P}(S) - v_\mathrm{P} \begin{cases} = O(\|S\|) & \text{if strong duality holds for (P);} \\ \text{can be "huge"} & \text{if } v_\mathrm{D} > v_\mathrm{P}; \\ = O(\|S\|^\gamma) \text{ for some fixed } \gamma \in (0,1) & \text{if strong duality fails for (P)} \\ & \text{and } v_\mathrm{P} = v_\mathrm{D}. \end{cases} \tag{7.2}$$

While the first two cases in (7.2) are relatively straightforward, the last case is far from obvious. The parameter γ can be expressed in terms of the degree of singularity of the LMI defining the feasible region of (P). This chapter aims at proving (7.2). We first review some asymptotic properties of SDP in Section 7.1. Then we provide a few illustrative examples in Section 7.2. Then we consider Case 1 in Section 7.3, Case 2 in Section 7.4, and Case 3 in Section 7.5. Section

7.5 requires some results concerning the degree of singularity; we provide those relevant results in Section 7.5.3.

We remark that there is quite a volume of results on sensitivity analysis on nonlinear programs. To mention a few, [45] performs sensitivity analysis on SDPs assuming that the Slater condition holds for both the primal and the dual. In [78], the optimal value function $\phi(u) := \inf_x F(x,u)$ is considered, and its directional derivative is studied. [14] studies the perturbation theory of nonlinear programs, in which Section 7.3 focuses on the case where the dual is not solvable.

7.1 Review: asymptotic properties of SDP

Before we proceed, we first review some basic definitions and results concerning the asymptotic feasibility and optimal value of SDP. This terminology is needed in the examples in Section 7.2.

A sequence $\{y^{(k)}\}_k$ is said to be *asymptotically feasible* for (P) if there exists a sequence $\{Z^{(k)}\}_k \subset \mathbb{S}^n_+$ such that $Z^{(k)} + \mathcal{A}^* y^{(k)} \to C$ as $k \to \infty$. We say that (P) is *weakly infeasible* if (P) is not feasible but possesses an asymptotically feasible sequence, and that (P) is *strongly infeasible* if (P) does not have an asymptotically feasible sequence. Similarly, a sequence $\{X^{(k)}\}_k$ is said to be asymptotically feasible for (D) if $X^{(k)} \succeq 0$ for all k and $\lim_k \mathcal{A}(X^{(k)}) = b$. Strong infeasibility and weak infeasibility of (D) are defined similarly as for (P).

Define the *asymptotic optimal value* of (P) as

$$v_{\mathrm{P}}^a := \sup \left\{ \limsup_k b^\top y^{(k)} : \left\{y^{(k)}\right\}_k \text{ is asymptotically feasible for (P)} \right\}, \tag{7.3}$$

and the asymptotic optimal value of (D) as

$$v_{\mathrm{D}}^a := \inf \left\{ \liminf_k \langle C, X^{(k)} \rangle : \left\{X^{(k)}\right\}_k \text{ is asymptotically feasible for (D)} \right\}.$$

We take the convention that $v_{\mathrm{P}}^a = -\infty$ (respectively, $v_{\mathrm{D}}^a = +\infty$) if (P) (respectively, (D)) is strongly infeasible. Note that if (P) is feasible, then $v_{\mathrm{P}}^a \geq v_{\mathrm{P}}$. As we can see in Example 7.2.2 below, strict inequality may hold.

We say that $\hat{y} \in \mathbb{R}^m$ is an *improving direction* for (P) if $-\mathcal{A}^*\hat{y} \succeq 0$ and $b^\top \hat{y} \geq 1$, and that $\{y^{(k)}\}_k \subset \mathbb{R}^m$ is an *improving direction sequence* for (P) if there exists a sequence $\{Z^{(k)}\}_k \subset \mathbb{S}^n_+$ such that $Z^{(k)} + \mathcal{A}^* y^{(k)} \to 0$ and $b^\top y^{(k)} \geq 1$ for all k. Improving direction sequences and improving directions for (P), respectively, serve as certificates of infeasibility and strong infeasibility of the dual (D):

Lemma 7.1.1 ([62], Lemmas 5 and 6)**.** *The SDP* (D) *is infeasible if and only if* (P) *possesses an improving direction sequence.* (D) *is strongly infeasible if and only if* (P) *possesses an improving direction.*

The dual of an SDP satisfying Assumption 7.1 cannot be strongly infeasible:

Theorem 7.1.2 ([35])**.** *If* (P) *is feasible and* $v_{\mathrm{P}} < +\infty$*, then* (D) *is either feasible or weakly infeasible, and* $v_{\mathrm{D}}^a = v_{\mathrm{P}}$*.*

If both (P) and (D) are feasible, then *weak duality*, i.e., $v_{\mathrm{P}} \leq v_{\mathrm{D}}$, implies that both (P) and (D) have finite optimal value, and Theorem 7.1.2 implies that

$$v_{\mathrm{P}}^a = v_{\mathrm{D}} \geq v_{\mathrm{P}} = v_{\mathrm{D}}^a. \tag{7.4}$$

7.2 Examples

In this section we give some examples of SDPs where strong duality fails, and we examine some possible perturbations that could lead to a *big* change in the optimal value. By abuse of notation, in this section we restrict the function $\mathrm{val}_{\mathrm{P}}(\cdot)$ defined in (7.1) on a fixed direction $\epsilon \mapsto \epsilon \hat{S}$ for some fixed $\hat{S}$, so $\mathrm{val}_{\mathrm{P}}(\cdot)$ is a function on $\mathbb{R}_+$.

Example 7.2.1. [90] (D) **is infeasible.** *For* $\epsilon \geq 0$*, consider*

$$\mathrm{val}_{\mathrm{P}}(\epsilon) := \sup \left\{ y_2 : \begin{pmatrix} y_1 & y_2 & y_3 \\ y_2 & y_3 & 0 \\ y_3 & 0 & 0 \end{pmatrix} \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \epsilon \end{pmatrix} \right\}, \tag{7.5}$$

$$\mathrm{val}_{\mathrm{D}}(\epsilon) := \inf \left\{ \epsilon X_{33} : X_{11} = 0,\ X_{12} = 1,\ 2X_{13} + X_{22} = 0,\ X \succeq 0 \right\}. \tag{7.6}$$

The dual (7.6) *is infeasible. When* $\epsilon = 0$*, the primal* (7.5) *has optimal value* $\mathrm{val}_{\mathrm{P}}(0) = 0$ *while the asymptotic optimal value is* $+\infty$*. Indeed, consider for* $k \in \mathbb{N}$*,*

$$Z^{(k)} = \begin{pmatrix} k^2 & -k & 1 \\ -k & 1 & 0 \\ 1 & 0 & \frac{1}{k} \end{pmatrix},\ y^{(k)} = \begin{pmatrix} -k^2 \\ k \\ 1 \end{pmatrix}.$$

Then

$$Z^{(k)} + \mathcal{A}^* y^{(k)} = \begin{pmatrix} k^2 & -k & 1 \\ -k & 1 & 0 \\ 1 & 0 & \frac{1}{k} \end{pmatrix} + \begin{pmatrix} -k^2 & k & -1 \\ k & -1 & 0 \\ -1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{k} \end{pmatrix} \to 0,$$

(meaning that $\left\{y^{(k)}\right\}_k$ *is asymptotically feasible) and* $b^\top y^{(k)} = k \to \infty$ *as* $k \to \infty$*. Hence* $v_{\mathrm{P}}^a(0) = +\infty > 0 = \mathrm{val}_{\mathrm{P}}(0)$*.*

Now we show that for any $\epsilon > 0$*, we have* $\mathrm{val}_{\mathrm{P}}(\epsilon) = +\infty$*. For all sufficiently large* $k \in \mathbb{R}$*,*

$$\begin{pmatrix} k^2 & -k & 1 \\ -k & 1 & 0 \\ 1 & 0 & \epsilon \end{pmatrix} \succeq 0,$$

so $\begin{pmatrix} -k^2 \\ k \\ -1 \end{pmatrix}$ *is feasible for the perturbed problem* (7.5)*. Hence* $\mathrm{val}_{\mathrm{P}}(\epsilon) = +\infty$*.*

Example 7.2.2. [73] nonzero finite duality gap. *Fix any* $\alpha > 0$*. For* $\epsilon \geq 0$*, consider the primal-dual pair*

$$\mathrm{val}_{\mathrm{P}}(\epsilon) := \sup_y \left\{ y_2 : \begin{pmatrix} y_2 & 0 & 0 \\ 0 & y_1 & y_2 \\ 0 & y_2 & 0 \end{pmatrix} \preceq \begin{pmatrix} \alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \epsilon \end{pmatrix} \right\}, \tag{7.7}$$

$$\mathrm{val}_{\mathrm{D}}(\epsilon) := \inf_X \left\{ \alpha X_{11} + \epsilon X_{33} : X_{22} = 0,\ X_{11} + X_{23} = 1,\ X \succeq 0 \right\}. \tag{7.8}$$

Let $\mathcal{F}_{\mathrm{P}}(\epsilon)$ *be the set of feasible solutions* y *for* (7.7) *and* $\mathcal{F}_{\mathrm{D}}(\epsilon)$ *be the set of feasible solutions* X *for* (7.8)*. For* $\epsilon = 0$*, we get*

$$\mathcal{F}_{\mathrm{P}}(0) = \mathbb{R}_- \times 0,\ \mathcal{F}_{\mathrm{P}}^Z(0) = \left\{ \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \gamma & 0 \\ 0 & 0 & 0 \end{pmatrix} : \gamma \geq 0 \right\},\ \mathcal{F}_{\mathrm{D}}(0) = \left\{ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \beta \end{pmatrix} : \beta \geq 0 \right\}.$$

So $\mathrm{val}_{\mathrm{P}}(0) = 0 < \alpha = \mathrm{val}_{\mathrm{D}}(0) = v_{\mathrm{P}}^a(0)$*. (To see that* $v_{\mathrm{P}}^a = \alpha$*, consider the sequences* $y^{(k)} = \begin{pmatrix} -\alpha k^2 \\ \alpha \end{pmatrix} \in \mathbb{R}^2$ *and* $Z^{(k)} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & \alpha^2 k & -\alpha \\ 0 & -\alpha & \frac{1}{k} \end{pmatrix} \in \mathbb{S}_+^3$*. We have* $Z^{(k)} + \mathcal{A}^* y^{(k)} = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{1}{k} \end{pmatrix}$ *and* $b^\top y^{(k)} = \alpha$ *for all* k*.)*

Now consider $\epsilon > 0$*. Then*

$$\begin{pmatrix} \alpha - y_2 & 0 & 0 \\ 0 & -y_1 & -y_2 \\ 0 & -y_2 & \epsilon \end{pmatrix} \succeq 0$$

if and only if $y_2 \leq \alpha$ *and* $y_1 \leq -|y_2|/\epsilon$*. So* $\mathrm{val}_{\mathrm{P}}(\epsilon) = \alpha = v_{\mathrm{P}}^a(0)$*. On the other hand, the objective of the dual becomes* $\alpha X_{11} + \epsilon X_{33}$ *and the constraints are unchanged. Hence* $\mathrm{val}_{\mathrm{D}}(\epsilon) = \alpha = \mathrm{val}_{\mathrm{D}}(0)$*, and* $\mathrm{val}_{\mathrm{P}}(\epsilon) - \mathrm{val}_{\mathrm{P}}(0) = \alpha$*.*

(Observe that the primal requires only one iteration of facial reduction to identify the minimal face, but then the dual of the reduced primal would still fail the Slater condition.)

Example 7.2.3. [69] zero duality gap but v_{D} is unattained. *Consider*

$$\mathrm{val_P}(0) := \sup\left\{2y_1 : y_1\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \preceq \begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\right\}, \tag{7.9}$$

$$\mathrm{val_D}(0) := \inf\left\{X_{11} : \begin{pmatrix}X_{11} & 1\\ 1 & X_{22}\end{pmatrix} \succeq 0\right\}. \tag{7.10}$$

On the one hand, $y = 0$ is the only feasible solution for (7.9) so $\mathrm{val_P}(0) = 0$. On the other hand, $\mathrm{val_D}(0) = 0$ but is not attained. Hence strong duality does not hold for (7.9).

Now consider

$$\mathrm{val_P}(\epsilon) := \sup\left\{2y_1 : y_1\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix} \preceq \begin{pmatrix}1 & 0\\ 0 & \epsilon\end{pmatrix}\right\}, \tag{7.11}$$

$$\mathrm{val_D}(\epsilon) := \inf\left\{X_{11} + \epsilon X_{22} : \begin{pmatrix}X_{11} & 1\\ 1 & X_{22}\end{pmatrix} \succeq 0\right\}, \tag{7.12}$$

with $\epsilon > 0$. Since (7.11) satisfies the Slater condition and y_1 is feasible for (7.11) if and only if $|y_1| \leq \sqrt{\epsilon}$, we have $\mathrm{val_P}(\epsilon) = \mathrm{val_D}(\epsilon) = 2\sqrt{\epsilon}$ (and $\mathrm{val_D}(\epsilon)$ is attained), and $\mathrm{val_P}(\epsilon) - \mathrm{val_P}(0) = 2\sqrt{\epsilon}$.

Example 7.2.4. Zero duality gap but v_{D} is unattained. *This example generalizes Example 7.2.3. We consider an SDP on on $\mathbb{S}^n_+$ that requires $n-1$ iterations of facial reduction on (7.13) to identify the minimal face of $\mathbb{S}^n_+$ containing its feasible region. We show that there exists a feasible perturbation S such that $\mathrm{val_P}(S) - \mathrm{val_P}(0) = 2\|S\|^{1/2^{n-1}}$.*

Let $n \geq 3$ and $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^{n-1}$ be defined by the matrices

$$A_1 = e_1e_2^\top + e_2e_1^\top,\ A_k = e_ke_k^\top + e_1e_{k+1}^\top + e_{k+1}e_1^\top,\ \forall\, k \in 2 : n-1.$$

Consider

$$\mathrm{val_P}(0) = \sup\left\{2y_1 : Z = e_1e_1^\top - \mathcal{A}^*y \succeq 0\right\}. \tag{7.13}$$

Since $Z = e_1e_1^\top - \mathcal{A}^*y$ always have $Z_{nn} = 0$, $Z \succeq 0$ if and only if $Z = e_1e_1^\top$, i.e., $y = 0$. Hence $\mathrm{val_P}(0) = 0$. On the other hand, the dual

$$\begin{aligned}\mathrm{val_D}(0) = \quad \inf\{&X_{11} : X_{12} = 1,\ X_{22} + 2X_{13} = 0,\ X_{33} + 2X_{14} = 0,\ \ldots,\\ &X_{n-1,n-1} + 2X_{1n} = 0,\ X \succeq 0\}.\end{aligned} \tag{7.14}$$

has an optimal value $\mathrm{val_D}(0) = 0$ but is not attained. Hence strong duality does not hold for (7.13).

Now for $\epsilon > 0$ consider

$$\mathrm{val_P}(\epsilon) := \sup\left\{2y_1 : \mathcal{A}^* y \preceq e_1 e_1^\top + \epsilon e_n e_n^\top\right\}, \tag{7.15}$$

$$\mathrm{val_D}(\epsilon) := \inf\left\{X_{11} + \epsilon X_{nn} : \mathcal{A}(X) = b,\ X \succeq 0\right\}. \tag{7.16}$$

It is not difficult to see that (7.15) *satisfies the Slater condition: suppose $D \succeq 0$ satisfies $\langle C, D\rangle = 0$ and $\mathcal{A}(D) = 0$. It suffices to show that $D = 0$ (see Theorem 3.3.10). Indeed, $D \succeq 0$ and $\langle C, D\rangle = 0$ imply $D_{11} = D_{nn} = 0$. But this in turn implies that $D_{n-1,n-1} = \cdots = D_{22} = 0$. Hence $D = 0$.*

Now note that y is feasible for (7.15) *if and only if*

$$0 \geq y_{n-1} \geq -\epsilon^{1/2}, \quad 0 \geq y_{n-2} \geq -\epsilon^{1/4}, \ \ldots, \quad 0 \geq y_2 \geq -\epsilon^{1/2^{n-2}}, \quad |y_1| \leq \epsilon^{1/2^{n-1}}.$$

Hence $\mathrm{val_P}(\epsilon) = \mathrm{val_D}(\epsilon) = 2\epsilon^{1/2^{n-1}}$, and $\mathrm{val_P}(\epsilon) - \mathrm{val_P}(0) = 2\epsilon^{1/2^{n-1}}$.

Example 7.2.5. Zero duality gap but v_D is unattained. *Let*

$$A_1 = -E_{11},\ A_2 = -E_{22}, A_3 = e_3 e_4^\top + e_4 e_3^\top,\ A_4 = e_1 e_3^\top + e_3 e_1^\top,\ C = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

and $b = \begin{pmatrix} 0 & -1 & 2 & 0 \end{pmatrix}^\top$. Then (P) *reads*

$$\sup\left\{-y_2 + 2y_3 : \begin{pmatrix} y_1 & 1 & -y_4 & 0 \\ 1 & y_2 & 0 & 0 \\ -y_4 & 0 & 0 & -y_3 \\ 0 & 0 & -y_3 & 1 \end{pmatrix} \succeq 0\right\} = 0, \tag{7.17}$$

which is unattained. (Note that y feasible must satisfy $y_3 = y_4 = 0$.) Meanwhile, (D) *reads*

$$\begin{aligned} &\inf\left\{2X_{12} + X_{44} : X_{11} = 0,\ X_{22} = 1,\ X_{34} = 1,\ X_{13} = 0,\ X \succeq 0\right\} \\ =&\inf\left\{X_{44} : X = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & * & * \\ 0 & * & * & 1 \\ 0 & * & 1 & * \end{pmatrix} \succeq 0\right\} = 0, \end{aligned}$$

which is unattained too. Hence strong duality does not hold for (7.17). *Now for any $\epsilon > 0$,*

$$\mathrm{val_P}(\epsilon) := \sup\left\{-y_2 + 2y_3 : \begin{pmatrix} y_1 & 1 & -y_4 & 0 \\ 1 & y_2 & 0 & 0 \\ -y_4 & 0 & \epsilon & -y_3 \\ 0 & 0 & -y_3 & 1 \end{pmatrix} \succeq 0\right\} = 2\sqrt{\epsilon}.$$

In other words, $\mathrm{val_P}(\epsilon) - \mathrm{val_P}(0) = 2\sqrt{\epsilon}$.

7.3 Case 1: strong duality holds for (P)

If strong duality holds for (P), then a small perturbation leads to little change in the optimal value:

Theorem 7.3.1 ([17, 83])**.** *The value function* $\mathrm{val_P} : \mathbb{S}^n \to [-\infty, \infty]$ *is concave. Moreover, the equality* $v_\mathrm{P} = v_\mathrm{D}$ *holds if and only if the* $\mathrm{val_P}(0)$ *is upper semicontinuous at* 0*. In such case,* X^* *is an optimal solution of* (D) *if and only if* $\mathrm{val_P}(S) - \mathrm{val_P}(0) \leq \langle X^*, S\rangle$ *for all* $S \in \mathbb{S}^n$*.*

As a corollary we get the following result.

Corollary 7.3.2. *Suppose that strong duality holds for* (P)*. Then there exists a constant* $\kappa > 0$ *such that for any* $S \in \mathbb{S}^n$ *with* (7.1) *feasible,* $\mathrm{val_P}(S) - \mathrm{val_P}(0) \leq \kappa\|S\|$*.*

What if strong duality does not hold? We consider the two different cases: (1) (P)-(D) has a nonzero duality gap, and (2) the duality gap is zero, i.e., $v_\mathrm{P} = v_\mathrm{D}$, but v_D is unattained.

7.4 Case 2: nonzero duality gap

We already saw in Section 7.1 that if both (P) and (D) are feasible, then $v_\mathrm{D} = v_\mathrm{P}^a$, so the duality gap is given by $v_\mathrm{P}^a - v_\mathrm{P}$. We first show that $v_\mathrm{P}^a - v_\mathrm{P}$ is the duality gap between (P) and (D) even when (D) is infeasible, as a result of Lemma 7.4.1 below. In particular, (7.4) holds whenever (P) satisfies Assumption 7.1, and (P)-(D) having a nonzero duality gap is equivalent to $v_\mathrm{P} < v_\mathrm{P}^a \in \mathbb{R} \cup \{+\infty\}$. We show in Proposition 7.4.2 that in such case there exists an arbitrarily small right-hand side perturbation such that the new optimal value jumps by at least $v_\mathrm{P}^a - v_\mathrm{P}$ (which could be $+\infty$).

Lemma 7.4.1. *Suppose that* (P) *satisfies Assumption 7.1 but its dual* (D) *is infeasible. Then* $v_\mathrm{P}^a = +\infty$*.*

Proof. Since (D) is infeasible, by Lemma 7.1.1 there exists a sequence $\left\{(y^{(k)}, Z^{(k)})\right\}_k$ satisfying

$$b^\top y^{(k)} \geq 1 \text{ and } Z^{(k)} \succeq 0 \text{ for all } k, \quad \text{and} \quad \lim_k \left(Z^{(k)} + \mathcal{A}^* y^{(k)}\right) = 0.$$

By Assumption 7.1 and Theorem 7.1.2, (D) cannot be strongly infeasible. Thus $Z^{(k)} + \mathcal{A}^* y^{(k)} \neq 0$ for all k. (Otherwise $y^{(k)}$ would be an improving direction for (P), implying that (D) is strongly infeasible from Lemma 7.1.1.)

For each k, define

$$\hat{y}^{(k)} := \frac{1}{\|Z^{(k)} + \mathcal{A}^* y^{(k)}\|^{1/2}} y^{(k)}, \quad \hat{Z}^{(k)} := \frac{1}{\|Z^{(k)} + \mathcal{A}^* y^{(k)}\|^{1/2}} Z^{(k)}.$$

Then $\hat{Z}^{(k)} \succeq 0$ for all k,

$$\|\hat{Z}^{(k)} + \mathcal{A}^* \hat{y}^{(k)}\| = \|Z^{(k)} + \mathcal{A}^* y^{(k)}\|^{1/2} \implies \lim_k \left(\hat{Z}^{(k)} + \mathcal{A}^* \hat{y}^{(k)}\right) = 0$$

and

$$b^\top \hat{y}^{(k)} \geq \frac{1}{\|Z^{(k)} + \mathcal{A}^* y^{(k)}\|^{1/2}} \to +\infty \implies \lim_k b^\top \hat{y}^{(k)} = +\infty.$$

On the other hand, since (P) is feasible, let $\hat{y} \in \mathbb{R}^m$ satisfy $C - \mathcal{A}^* \hat{y} \succeq 0$. Then $\left\{\hat{y} + \hat{y}^{(k)}\right\}_k$ is asymptotically feasible for (P), and $\lim_k b^\top (\hat{y} + \hat{y}^{(k)}) = +\infty$. Hence $v_{\mathrm{P}}^a = +\infty$. □

Proposition 7.4.2. *Suppose that* (P) *satisfies Assumption 7.1. Then for every $\epsilon > 0$, there exists $S \in \mathbb{S}^n$ such that*

(1) $\|S\| \leq \epsilon$,

(2) the perturbed problem (7.1) *is strictly feasible, and*

(3) the optimal value $\mathrm{val}_{\mathrm{P}}(S)$ *of the perturbed problem* (7.1) *is no less than the asymptotic optimal value* v_{P}^a *of* (P)*, defined in* (7.3).

Proof. Let sequences $\left\{y^{(k)}\right\}_k$ and $\left\{Z^{(k)}\right\}_k$ satisfy

$$Z^{(k)} \succeq 0 \text{ for all } k, \quad \lim_k \left(Z^{(k)} + \mathcal{A}^* y^{(k)}\right) = C, \quad \text{and} \quad \lim_k b^\top y^{(k)} = v_{\mathrm{P}}^a.$$

Fix any $\epsilon > 0$. There exists $n_0 \in \mathbb{N}$ such that $\|\mathcal{A}^*(y^{(k)} - y^{(l)}) + Z^{(k)} - Z^{(l)}\| \leq \frac{1}{4}\epsilon$ and $\|C - Z^{(k)} - \mathcal{A}^* y^{(k)}\| \leq \frac{1}{2}\epsilon$ for all $k, l \geq k_0$. In particular, $\mathcal{A}^*(y^{(k)} - y^{(l)}) + Z^{(k)} \succeq Z^{(l)} - \frac{\epsilon}{4} I$, and $S := C - (Z^{(k_0)} + \mathcal{A}^* y^{(k_0)} + \frac{\epsilon}{2} I)$ satisfies $\|S\| \leq \epsilon$. Moreover, for all $k \geq k_0$,

$$C - \mathcal{A}^* y^{(k)} = \mathcal{A}^*(y^{(k_0)} - y^{(k)}) + Z^{(k_0)} + \frac{\epsilon}{2} I + S \succeq Z^{(k)} - \frac{\epsilon}{4} I + \frac{\epsilon}{2} I + S \succ S,$$

showing that $y^{(k)}$ is a (strictly) feasible solution of (7.1). Hence, $\mathrm{val}_{\mathrm{P}}(S) \geq b^\top y^{(k)}$ for all $k \geq k_0$. Taking $k \to \infty$, we get $\mathrm{val}_{\mathrm{P}}(S) \geq v_{\mathrm{P}}^a$. □

A direct result of Lemma 7.4.1 and Proposition 7.4.2 is that there exists $S \in \mathbb{S}^n$ of arbitrarily small norm such that (7.1) is feasible and the jump in optimal value $\mathrm{val}_{\mathrm{P}}(S) - \mathrm{val}_{\mathrm{P}}(0)$ is no less than the duality gap $v_{\mathrm{D}} - v_{\mathrm{P}}$.

Theorem 7.4.3. *Suppose that* (P) *satisfies Assumption 7.1 and* $v_{\mathrm{P}} < v_{\mathrm{D}} \in \mathbb{R} \cup \{+\infty\}$. *Then for every* $\epsilon > 0$, *there exists* $S \in \mathbb{S}^n$ *such that*

(1) $\|S\| \leq \epsilon$,

(2) the perturbed problem (7.1) *is strictly feasible, and*

(3) $\mathrm{val}_{\mathrm{P}}(S) - \mathrm{val}_{\mathrm{P}}(0) \geq v_{\mathrm{D}} - v_{\mathrm{P}} > 0$.

Proof. If (D) is feasible, then $v_{\mathrm{D}} > v_{\mathrm{P}} > -\infty$ so $v_{\mathrm{D}} = v_{\mathrm{P}}^a$, by Lemma 7.1.1. If (D) is infeasible, then by Lemma 7.4.1 $v_{\mathrm{D}} = +\infty = v_{\mathrm{P}}^a$. Hence by Proposition 7.4.2 for every $\epsilon > 0$ there exists S satisfying (1), (2) and

$$\mathrm{val}_{\mathrm{P}}(S) - \mathrm{val}_{\mathrm{P}}(0) \geq v_{\mathrm{P}}^a - v_{\mathrm{P}} = v_{\mathrm{D}} - v_{\mathrm{P}}.$$

□

7.5 Case 3: strong duality fails but duality gap is zero

If the duality gap between (P) and (D) is zero and yet strong duality fails, then by Theorem 7.3.1, the function $\mathrm{val}_{\mathrm{P}}(\cdot)$ is upper semicontinuous at 0 but $\partial(-\mathrm{val}_{\mathrm{P}}(\cdot))(0) = \varnothing$. The results in this section rely on the following error bound result for linear matrix inequalities (LMI).

Theorem 7.5.1 ([86], Theorem 3.3)**.** *Suppose that the set*

$$\mathcal{F}_{\mathrm{P}}^Z := \left\{ Z \in \mathbb{S}_+^n : Z = C - \mathcal{A}^* y \text{ for some } y \in \mathbb{R}^m \right\}$$

is nonempty. Then there exist constants $\kappa > 0$ *and* $\bar{\epsilon} \in (0, 1)$ *such that for any* $\epsilon \in (0, \bar{\epsilon})$ *and any* $\tilde{Z} \in \mathbb{S}^n$ *satisfying*

$$\mathrm{dist}(\tilde{Z}, C + \mathrm{range}(\mathcal{A}^*)) \leq \epsilon,\ \lambda_{\min}(\tilde{Z}) \geq -\epsilon,$$

we have

$$\mathrm{dist}(\tilde{Z}, \mathcal{F}_{\mathrm{P}}^Z) \leq \kappa(1 + \|\tilde{Z}\|)\epsilon^{1/2^{\mathrm{d}(\mathcal{A},C)}},$$

where $\mathrm{d}(\mathcal{A}, C)$ *is the degree of singularity of the linear subspace* $\mathrm{range}(\mathcal{A}_C^*)$, *defined in Definition 7.5.5 and* (7.30).

We will use Theorem 7.5.1 to show that if $S \in \mathbb{S}^n$ is such that $\|S\|$ is sufficiently small and (7.1) is feasible, then $\mathrm{val}_{\mathrm{P}}(S) - \mathrm{val}_{\mathrm{P}}(0) = O(\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}})$. We first deal with the case where (D) satisfies the Slater condition, in Section 7.5.1; then we use this to prove the general result in Section 7.5.2.

7.5.1 Case 3(a): (D) satisfies the Slater condition

We first prove a weaker result assuming that the dual (D) satisfies the Slater condition.

Proposition 7.5.2. *Suppose that* (P) *satisfies Assumption 7.1 and that* (D) *satisfies the Slater condition. Then there exist constants* $\kappa > 0$ *and* $\bar{\epsilon} \in (0,1)$ *such that for any* $S \in \mathbb{S}^n$ *with* $0 < \|S\| \le \bar{\epsilon}$ *and* (7.1) *feasible,*

$$\mathrm{val_P}(S) - \mathrm{val_P}(0) \le \kappa \|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}, \tag{7.18}$$

where $\mathrm{d}(\mathcal{A},C)$ *denotes the degree of singularity of the linear subspace* $\mathrm{range}(\mathcal{A}_C^*)$*, defined in Definition 7.5.5.*

Proof. Let $\kappa_0 > 0$ and $\bar{\epsilon} \in (0,1)$ be such that for any $Y \in \mathbb{S}^n$ and $\epsilon \in (0,\bar{\epsilon})$ with

$$\mathrm{dist}(Y, C + \mathrm{range}(\mathcal{A}^*)) \le \epsilon,\ \lambda_{\min}(Y) \ge -\epsilon,$$

we get

$$\mathrm{dist}(Y, \mathcal{F}_{\mathrm{P}}^Z) \le \kappa_0(1 + \|Y\|)\epsilon^{1/2^{\mathrm{d}(\mathcal{A},C)}}. \tag{7.19}$$

Let $\tilde{X}$ be a strictly feasible solution of (D). Fix any $S \in \mathbb{S}^n$ with (7.1) feasible and $\|S\| \le \bar{\epsilon}$. Since the dual of (7.1) has a (strictly) feasible solution $\tilde{X}$, we get $\mathrm{val_P}(S) < +\infty$.

Fix any $\delta \in (0,1)$. Then there exist $\tilde{y}, \tilde{Z}$ satisfying

$$b^\top \tilde{y} \ge \mathrm{val_P}(S) - \delta, \quad \tilde{Z} = C - S - \mathcal{A}^*\tilde{y} \succeq 0.$$

For any $y \in \mathbb{R}^m$ satisfying $Z := C - \mathcal{A}^* y \succeq 0$,

$$\begin{aligned}\mathrm{val_P}(S) - \mathrm{val_P}(0) &\le b^\top \tilde{y} - b^\top y + \delta \\ &= \langle \tilde{X}, C - S - \tilde{Z}\rangle - \langle \tilde{X}, C - Z\rangle + \delta \\ &\le \|\tilde{X}\|\|Z - (\tilde{Z} + S)\| + \delta.\end{aligned}$$

Minimizing over all $Z \in \mathcal{F}_{\mathrm{P}}^Z$,

$$\mathrm{val_P}(S) - \mathrm{val_P}(0) \le \|\tilde{X}\|\,\mathrm{dist}(\tilde{Z} + S,\, \mathcal{F}_{\mathrm{P}}^Z) + \delta. \tag{7.20}$$

But $\tilde{Z} + S \in C + \mathrm{range}(\mathcal{A}^*)$ and $\tilde{Z} + S \succeq S \succeq -\|S\| I$, implying $\lambda_{\min}(\tilde{Z} + S) \ge -\|S\|$. Hence by (7.19), $\mathrm{dist}(\tilde{Z} + S,\, \mathcal{F}_{\mathrm{P}}^Z) \le \kappa_0(1 + \|\tilde{Z} + S\|)\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}$. Hence by (7.20),

$$\mathrm{val_P}(S) - \mathrm{val_P}(0) \le \kappa_0\|\tilde{X}\|(1 + \|\tilde{Z} + S\|)\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}} + \delta. \tag{7.21}$$

Now we estimate $\|\tilde{Z}+S\|$. Observe that

$$\lambda_{\min}(\tilde{X})\|\tilde{Z}\| \leq \langle C-S-\mathcal{A}^*\tilde{y}, \tilde{X}\rangle \leq \langle C-S, \tilde{X}\rangle - \mathrm{val_P}(S) + \delta. \tag{7.22}$$

Using (7.22) and the assumption that $\|S\| < 1$, the right hand side of (7.21) no greater than the expression

$$\begin{aligned}
&\kappa_0\|\tilde{X}\|\left(1+\|S\|+\frac{1}{\lambda_{\min}(\tilde{X})}\left(\langle C-S,\tilde{X}\rangle - \mathrm{val_P}(S)+\delta\right)\right)\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}+\delta \\
\leq\ &\kappa_0\frac{\|\tilde{X}\|}{\lambda_{\min}(\tilde{X})}\left(2\lambda_{\min}(\tilde{X})+\langle C-S,\tilde{X}\rangle - \mathrm{val_P}(S)+\delta\right)\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}+\delta \\
\leq\ &\kappa_0\frac{\|\tilde{X}\|}{\lambda_{\min}(\tilde{X})}\left(2\lambda_{\min}(\tilde{X})+\langle C-S,\tilde{X}\rangle - \mathrm{val_P}(S)\right)\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}+\left(1+\kappa_0\frac{\|\tilde{X}\|}{\lambda_{\min}(\tilde{X})}\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}\right)\delta.
\end{aligned}$$

Putting back into (7.21), we get

$$\begin{aligned}
&\left(1+\kappa_0\frac{\|\tilde{X}\|}{\lambda_{\min}(\|\tilde{X}\|)}\|S\|^{1/2^{d(\mathcal{A},C)}}\right)(\mathrm{val_P}(S)-\mathrm{val_P}(0)) \\
\leq\ &\kappa_0\frac{\|\tilde{X}\|}{\lambda_{\min}(\tilde{X})}\left(2\lambda_{\min}(\tilde{X})+\langle C-S,\tilde{X}\rangle - \mathrm{val_P}(0)\right)\|S\|^{1/2^{d(\mathcal{A},C)}}+\left(1+\kappa_0\frac{\|\tilde{X}\|}{\lambda_{\min}(\tilde{X})}\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}\right)\delta.
\end{aligned} \tag{7.23}$$

Note that $2\lambda_{\min}(\tilde{X})+\langle C-S,\tilde{X}\rangle - \mathrm{val_P}(0) \leq 2\lambda_{\min}(\tilde{X})+\|\tilde{X}\|+\langle C,\tilde{X}\rangle - \mathrm{val_P}(0)$, which is positive by weak duality. Taking $\delta \searrow 0$ and defining

$$\kappa = \kappa_0\frac{\|\tilde{X}\|}{\lambda_{\min}(\tilde{X})}\left(2\lambda_{\min}(\tilde{X})+\|\tilde{X}\|+\langle C,\tilde{X}\rangle - \mathrm{val_P}(0)\right),$$

we get from (7.23) that

$$\left(1+\kappa_0\frac{\|\tilde{X}\|}{\lambda_{\min}(\|\tilde{X}\|)}\|S\|^{1/2^{d(\mathcal{A},C)}}\right)(\mathrm{val_P}(S)-\mathrm{val_P}(0)) \leq \kappa\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}},$$

so $\mathrm{val_P}(S)-\mathrm{val_P}(0) \leq \kappa\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}$. □

7.5.2 Case 3(b): (D) does not satisfy the Slater condition

Now we consider the case where $v_{\mathrm{P}} = v_{\mathrm{D}} \in \mathbb{R}$ but v_{D} is unattained and (D) fails the Slater condition. Such scenario can occur, as we can see in Example 7.2.5. We show that a bound of the form $\mathrm{val_P}(S)-\mathrm{val_P}(0) \leq \kappa\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}$ holds even in this case. The proof idea is to restrict (D) on its minimal face, and using the fact that such restriction does not change the degree of singularity of (P):

Lemma 7.5.3. *Suppose that* (P) *and* (D) *are feasible, and the minimal face of* (D) *is given by* $\tilde{P}\,\mathbb{S}^r_+\tilde{P}^\top$ *for some full column rank matrix* $\tilde{P} \in \mathbb{R}^{n\times r}$ *(with* $r > 0$*). Then*

$$\sup_y \left\{ b^\top y : \tilde{P}^\top (C - \mathcal{A}^* y)\tilde{P} \succeq 0 \right\}$$

is also feasible, and $\mathrm{d}(\mathcal{A}(\tilde{P}\cdot\tilde{P}^\top), \tilde{P}^\top C\tilde{P}) \le \mathrm{d}(\mathcal{A}, C)$.

The proof of Lemma 7.5.3 is given on Page 111 in Section 7.5.3. Now we prove the main results of this section.

Theorem 7.5.4. *Assume that* (P) *satisfies Assumption 7.1, that* $v_{\mathrm{P}} = v_{\mathrm{D}} \in \mathbb{R}$ *but* v_{D} *is unattained. Then there exist* $\bar{\epsilon} \in (0,1)$ *and* $\kappa > 0$ *such that for any* $S \in \mathbb{S}^n$ *with* (7.1) *feasible and* $\|S\| \le \bar{\epsilon}$*,*

$$\mathrm{val}_{\mathrm{P}}(S) - \mathrm{val}_{\mathrm{P}}(0) \le \kappa \|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}.$$

Proof. If (D) satisfies the Slater condition, then the statement in the theorem holds by Proposition 7.5.2. In the remainder of the proof we assume that (D) fails the Slater condition.

Since v_{D} is assumed to be unattained, the minimal face of (D) does not equal $\{0\}$.[1] Let $\tilde{P}\,\mathbb{S}^r_+\tilde{P}^\top$ be the minimal face of $\mathbb{S}^n_+$ containing the feasible region of (D) with $\tilde{P}^\top\tilde{P} = I$. Therefore we have $v_{\mathrm{D}} = \bar{v}_{\mathrm{D}}$, where

$$\bar{v}_{\mathrm{D}} := \inf_W \left\{ \langle C, \tilde{P}W\tilde{P}^\top\rangle : \mathcal{A}(\tilde{P}W\tilde{P}^\top) = b,\ W \succeq 0 \right\}. \tag{7.24}$$

By definition of minimal face, (7.24) satisfies the Slater condition. Note that since (D) has no optimal solution, (7.24) has no optimal solution either. The dual of (7.24) is given by

$$\bar{v}_{\mathrm{P}} := \sup_y \left\{ b^\top y : \tilde{P}^\top (C - \mathcal{A}^* y)\tilde{P} \succeq 0 \right\}, \tag{7.25}$$

Any y feasible for (P) is also feasible for (7.25). Hence we have

$$v_{\mathrm{P}} = v_{\mathrm{D}} = \bar{v}_{\mathrm{D}} \ge \bar{v}_{\mathrm{P}} \ge v_{\mathrm{P}}, \quad \text{i.e., } v_{\mathrm{P}} = \bar{v}_{\mathrm{P}}. \tag{7.26}$$

Moreover, the primal-dual pair (7.25)-(7.24) satisfies the assumptions in Proposition 7.4.2, which together with Lemma 7.5.3 implies that there exist constants $\kappa > 0$ and $\bar{\epsilon} \in (0,1)$ such that for any $\bar{S} \in \mathbb{S}^r$ with $0 < \|\bar{S}\| \le \bar{\epsilon}$ and (7.1) feasible,

$$\sup_y \left\{ b^\top y : \tilde{P}^\top (C - \mathcal{A}^* y)\tilde{P} \succeq \bar{S} \right\} - \bar{v}_{\mathrm{P}} \le \kappa \|\bar{S}\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}, \tag{7.27}$$

[1] If the minimal face of (D) is $\{0\}$, then $X = 0$ is the only feasible solution. This implies that $b = 0$ and $v_{\mathrm{D}} = 0$, so $v_{\mathrm{P}} = 0 = v_{\mathrm{D}}$ and any primal/dual feasible solution is optimal.

where $d(\mathcal{A}, C)$ denotes the degree of singularity of the linear subspace range$(\mathcal{A}_C^*)$.

Fix any $S \in \mathbb{S}^n$ with (7.1) feasible and $\|S\| \leq \bar{\epsilon}$. Then by weak duality and the fact that the feasible region of (D) is contained in $\tilde{P}\,\mathbb{S}_+^r\tilde{P}^\top$,

$$\begin{array}{rcl} \mathrm{val_P}(S) \leq & \inf_X \{\langle C - S, X\rangle : \mathcal{A}(X) = b,\ X \succeq 0\} \\ = & \inf_W \left\{\langle C - S, \tilde{P}W\tilde{P}^\top\rangle : \mathcal{A}(\tilde{P}W\tilde{P}^\top) = b,\ W \succeq 0\right\}, \end{array} \tag{7.28}$$

which satisfies the Slater condition. Since $\tilde{P}^\top(C - \mathcal{A}^*y)\tilde{P} \succeq \tilde{P}^\top S\tilde{P}$ is feasible, strong duality holds and

$$\inf_W \left\{\langle C - S, \tilde{P}W\tilde{P}^\top\rangle : \mathcal{A}(\tilde{P}W\tilde{P}^\top) = b,\ W \succeq 0\right\} = \sup_y \left\{b^\top y : \tilde{P}^\top(C - \mathcal{A}^*y)\tilde{P} \succeq \tilde{P}^\top S\tilde{P}\right\}. \tag{7.29}$$

Since $\|\tilde{P}^\top S\tilde{P}\| \leq \|S\| \leq \bar{\epsilon}$, we can use (7.27), (7.29) and (7.28) to get

$$\mathrm{val_P}(S) - \mathrm{val_P}(0) = \mathrm{val_P}(S) - \bar{v}_\mathrm{P} \leq \kappa\|\tilde{P}^\top S\tilde{P}\|^{1/2^{\mathrm{d}(\mathcal{A},C)}} \leq \kappa\|S\|^{1/2^{\mathrm{d}(\mathcal{A},C)}}.$$

□

Remark. The assumption that $v_\mathrm{P} = v_\mathrm{D}$ is important in this proof because this ensures that $v_\mathrm{P} = \bar{v}_\mathrm{P}$ in (7.26), which generally does not hold because the feasible region of (7.25) contains that of (P).

7.5.3 Facial reduction and degree of singularity

In this section, we provide supplementary results for Section 7.5.1 and 7.5.2.

We recall an equivalent form of Algorithm 4.1 from [86], outlined in Algorithm 7.1. The number of iterations of Algorithm 4.1 is called the *degree of singularity*. We formalize its definition below.

Definition 7.5.5. *Let $\bar{\mathcal{L}} \subseteq \mathbb{S}^n$ be a linear subspace. If $\bar{\mathcal{L}} \cap \mathbb{S}_+^n = \{0\}$ or $\bar{\mathcal{L}} \cap \mathbb{S}_{++}^n \neq \varnothing$, then the degree of singularity of $\bar{\mathcal{L}}$ is defined to be zero, i.e., $\mathrm{d}(\bar{\mathcal{L}}) := 0$. Otherwise, the degree of singularity of $\bar{\mathcal{L}}$ is defined as the length d of any sequence $D^{(0)}, D^{(1)}, \ldots, D^{(d-1)} \in \mathbb{S}^n$ such that*

1. $D^{(0)} \in \mathrm{ri}(\bar{\mathcal{L}}^\perp \cap \mathbb{S}_+^n)$,
2. $\mathrm{range}(D^{(0)}) = \mathrm{range}(Q_1^{(0)})$, $\ker(D^{(0)}) = \mathrm{range}(Q_2^{(0)})$,

   for some orthogonal matrix $Q^{(0)} = \begin{array}{c} \\ n \end{array} \begin{array}{c} \begin{array}{cc} n-n_1 & n_1 \end{array} \\ \left[\begin{array}{cc} Q_1^{(0)} & Q_2^{(0)} \end{array}\right] \end{array}$,

3. $\bar{\mathcal{L}}_1 = (Q^{(0)})^\top \bar{\mathcal{L}} Q^{(0)} = \begin{array}{c} \\ n-n_1 \\ n_1 \end{array} \begin{array}{c} \begin{array}{cc} n-n_1 & \quad n_1 \end{array} \\ \begin{bmatrix} (Q_1^{(0)})^\top \bar{\mathcal{L}} Q_1^{(0)} & (Q_1^{(0)})^\top \bar{\mathcal{L}} Q_2^{(0)} \\ (Q_2^{(0)})^\top \bar{\mathcal{L}} Q_1^{(0)} & (Q_2^{(0)})^\top \bar{\mathcal{L}} Q_2^{(0)} \end{bmatrix} \end{array}$,

4. *For* $i = 1 : d$,

$$D^{(i)} \in \operatorname{ri}\left(\bar{\mathcal{L}}_i^\perp \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{n_i} \end{bmatrix}^*\right) = \operatorname{ri}\left(\bar{\mathcal{L}}_i^\perp \cap \left\{X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} : X_{22} \succeq 0\right\}\right),$$

5. $D_{22}^{(d)} = 0$;

6. *for* $i = 1 : d-1$,

$$\begin{aligned} &\operatorname{range}(D_{22}^{(i)}) = \operatorname{range}(\bar{Q}_1^{(i)}), \quad \ker(D_{22}^{(i)}) = \operatorname{range}(\bar{Q}_2^{(i)}), \\ &\qquad\qquad \textit{for some orthogonal matrix } \begin{bmatrix} \bar{Q}_1^{((i))} & \bar{Q}_2^{((i))} \end{bmatrix}, \\ &Q_1^{(i)} := \begin{bmatrix} I & 0 \\ 0 & \bar{Q}_1^{(i)} \end{bmatrix}, \quad Q_2^{(i)} := \begin{bmatrix} 0 \\ \bar{Q}_2^{(i)} \end{bmatrix}, \quad Q^{(i)} := \begin{bmatrix} Q_1^{(i)} & Q_2^{(i)} \end{bmatrix}, \\ &\bar{\mathcal{L}}_{i+1} := (Q^{(i)})^\top \bar{\mathcal{L}}_i Q^{(i)}. \end{aligned}$$

Given a linear map $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ and $C \in \mathbb{S}^n$, we define

$$\mathrm{d}(\mathcal{A}, C) := \mathrm{d}\left(\{\mathcal{A}^* y + \alpha C : y \in \mathbb{R}^m,\ \alpha \in \mathbb{R}\}\right). \tag{7.30}$$

We first prove the simple fact that orthogonal transformations on linear subspaces do not change the degree of singularity.

Proposition 7.5.6. *Let* $U \in \mathbb{R}^{n\times n}$ *be an orthogonal matrix and* $\bar{\mathcal{L}} \subseteq \mathbb{S}^n$ *be a nonzero linear subspace. Then* $\mathrm{d}(U^\top \bar{\mathcal{L}} U) = \mathrm{d}(\bar{\mathcal{L}})$.

Proof. Since $U^\top \bar{\mathcal{L}} U \cap \mathbb{S}_+^n = U^\top (\bar{\mathcal{L}} \cap \mathbb{S}_+^n) U$, we have $\mathrm{d}(\bar{\mathcal{L}}) = 0$ if and only if $\mathrm{d}(U^\top \bar{\mathcal{L}} U) = 0$.

Suppose that $\mathrm{d}(\bar{\mathcal{L}}) > 0$. Then there exist $D^{(0)} \in \mathbb{S}^n$ and an orthogonal matrix $Q^{(0)} = \begin{array}{c} \\ n \end{array} \begin{array}{c} \begin{array}{cc} n-n_1 & \quad n_1 \end{array} \\ \begin{bmatrix} Q_1^{(0)} & Q_2^{(0)} \end{bmatrix} \end{array}$ such that

$$D^{(0)} \in \operatorname{ri}(\bar{\mathcal{L}}^\perp \cap \mathbb{S}_+^n), \quad \operatorname{range}(D^{(0)}) = \operatorname{range}(Q_1^{(0)}), \quad \ker(D^{(0)}) = \operatorname{range}(Q_2^{(0)}).$$

Since

$$\operatorname{ri}\left((U^\top \bar{\mathcal{L}} U)^\perp \cap \mathbb{S}_+^n\right) = \operatorname{ri}(U^\top \bar{\mathcal{L}}^\perp U \cap \mathbb{S}_+^n) = U^\top \operatorname{ri}(\bar{\mathcal{L}}^\perp \cap \mathbb{S}_+^n) U,$$

Algorithm 7.1: Sturm's procedure (for finding face$(\bar{\mathcal{L}} \cap \mathbb{S}^n_+, \mathbb{S}^n_+)$)

1 Input(*linear subspace* $\{0\} \neq \bar{\mathcal{L}}$ *of* $\mathbb{S}^n$);

(**1**) Let $D^{(0)} \in \mathrm{ri}(\bar{\mathcal{L}}^\perp \cap \mathbb{S}^n_+)$;

if $D^{(0)} = 0$ **or** $D^{(0)} \succ 0$ **then**

$\Longrightarrow$ $d(\bar{\mathcal{L}}) \leftarrow 0$;

else

$\Longrightarrow$ proceed to Step (**2**);

endif

(**2**) write $D^{(0)} = Q_1^{(0)} D_+^{(0)} (Q_1^{(0)})^\top$, where $Q^{(0)} = \begin{bmatrix} Q_1^{(0)} & Q_2^{(0)} \end{bmatrix}$ is orthogonal, $D_+^{(0)} \succ 0$;

$\bar{\mathcal{L}}_1 \leftarrow (Q^{(0)})^\top \bar{\mathcal{L}} Q^{(0)}$;

$n_0 \leftarrow n$, $n_1 \leftarrow$ # of columns of $Q_2^{(0)}$, $d \leftarrow 1$;

proceed to Step (**3**);

(**3**) let $D^{(d)} \in \mathrm{ri}\left(\bar{\mathcal{L}}_d^\perp \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}^{n_d}_+ \end{bmatrix}^*\right) = \mathrm{ri}\left(\bar{\mathcal{L}}_d^\perp \cap \left\{ X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} : X_{22} \succeq 0 \right\}\right)$;

if $D_{22}^{(d)} = 0$ **then**

$\Longrightarrow$ $d(\bar{\mathcal{L}}) \leftarrow d$;

else

$\Longrightarrow$ proceed to Step (**4**);

endif

(**4**) write $D_{22}^{(d)} = Q_1^{(d)} D_+^{(d)} (Q_1^{(d)})^\top$, where $Q^{(d)} = \begin{bmatrix} Q_1^{(d)} & Q_2^{(d)} \end{bmatrix}$ is orthogonal, $D_+^{(d)} \succ 0$;

define

$$\bar{Q}_1^{(d)} \leftarrow \begin{bmatrix} I & 0 \\ 0 & Q_1^{(d)} \end{bmatrix}, \quad \bar{Q}_2^{(d)} \leftarrow \begin{bmatrix} 0 \\ Q_2^{(d)} \end{bmatrix}, \quad \bar{Q}^{(d)} \leftarrow \begin{bmatrix} \bar{Q}_1^{(d)} & \bar{Q}_2^{(d)} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & Q^{(d)} \end{bmatrix};$$

$\bar{\mathcal{L}}_{d+1} \leftarrow (\bar{Q}^{(d)})^\top \bar{\mathcal{L}}_d \bar{Q}^{(d)}$;

$n_d \leftarrow$ # of columns of $Q_2^{(d)}$, $d \leftarrow d+1$;

return to Step (**3**).

---

we have $U^\top D^{(0)} U \in \mathrm{ri}\left((U^\top \bar{\mathcal{L}} U)^\perp \cap \mathbb{S}^n_+\right)$. Using

$$\mathrm{range}(U^\top D^{(0)} U) = \mathrm{range}(U^\top Q_1^{(0)}), \quad \ker(U^\top D^{(0)} U) = \mathrm{range}(U^\top Q_2^{(0)}),$$

and the fact that $\begin{bmatrix} U^\top Q_1^{(0)} & U^\top Q_2^{(0)} \end{bmatrix} = U^\top Q^{(0)}$ is orthogonal, the rotated linear subspace at Step 2 of Algorithm 7.1 is given by

$$(U^\top Q^{(0)})^\top (U^\top \bar{\mathcal{L}} U)(U^\top Q^{(0)}) = (Q^{(0)})^\top \bar{\mathcal{L}} Q^{(0)} = \bar{\mathcal{L}}_1.$$

Therefore, the remaining iterations of Algorithm 7.1 applied on $U^\top \bar{\mathcal{L}} U$ and on $\bar{\mathcal{L}}$ are the same. In particular, $\mathrm{d}(U^\top \bar{\mathcal{L}} U) = \mathrm{d}(\bar{\mathcal{L}})$. □

Let $P \in \mathbb{R}^{n\times r}$ be a full column rank matrix. The degree of singularity of the projection $P^\top \mathcal{L} P$ may be higher than that of $\mathcal{L}$. Consider

$$\mathcal{L} = \operatorname{span}\left\{ \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \right\}, \quad P = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \implies P^\top \mathcal{L} P = \operatorname{span}\left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \right\}.$$

Obviously, $\mathcal{L} \cap \mathbb{S}^3_+ = \{0\}$ so $\mathrm{d}(\mathcal{L}) = 0$. On the other hand, $P^\top \mathcal{L} P \cap \mathbb{S}^n_+$ is nonzero but has no interior point; in fact, $\mathrm{d}(P^\top \mathcal{L} P) = 1$.

On the other hand, we show that under some special condition (relevant to our discussion on facial reduction of (P)) the projection into a lower dimensional space would not increase the degree of singularity.

Theorem 7.5.7. *Suppose that* (P) *is feasible. Suppose that there exist* $V \in \mathbb{S}^n$ *and* $v \in \mathbb{R}^m$ *such that*

$$V = \begin{bmatrix} 0 & 0 \\ 0 & I_{n-r} \end{bmatrix} = \mathcal{A}^* v \succeq 0, \quad b^\top v, \quad 0 < r < n. \tag{7.31}$$

Let $\hat{C} = \begin{bmatrix} I_r & 0 \end{bmatrix} C \begin{bmatrix} I_r \\ 0 \end{bmatrix} \in \mathbb{S}^r$, $\hat{A}_i = \begin{bmatrix} I_r & 0 \end{bmatrix} A_i \begin{bmatrix} I_r \\ 0 \end{bmatrix} \in \mathbb{S}^r$ *for* $i \in 1:m$, *and define* $\hat{\mathcal{A}} : \mathbb{S}^r \to \mathbb{R}^m$ *using* $\hat{A}_1, \ldots, \hat{A}_m$. *Define* $\hat{\mathcal{L}} := \operatorname{span}(\hat{C}, \hat{A}_1, \ldots, \hat{A}_m)$. *Then* $\mathrm{d}(\hat{\mathcal{L}}) \le \mathrm{d}(\bar{\mathcal{L}})$.

Proof. Without loss of generality, assume that $d := \mathrm{d}(\hat{\mathcal{L}}) > 0$. Then there exist a sequence $\hat{D}^{(0)}, \hat{D}^{(1)}, \ldots, \hat{D}^{(d-1)}, \hat{D}^{(d)} \in \mathbb{S}^r$ and orthogonal matrices $\bar{Q}^{(0)} \in \mathbb{R}^{r\times r}, \bar{Q}^{(1)} \in \mathbb{R}^{r_1 \times r_1}, \ldots, \bar{Q}^{(d-1)} \in \mathbb{R}^{r_{d-1}\times r_{d-1}}$ such that:

1. $\hat{D}^{(0)} \in \mathrm{ri}(\hat{\mathcal{L}}^\perp \cap \mathbb{S}^r_+)$,
2. $\operatorname{range}(\hat{D}^{(0)}) = \operatorname{range}(\hat{Q}_1^{(0)})$, $\ker(\hat{D}^{(0)}) = \operatorname{range}(\hat{Q}_2^{(0)})$, $\hat{Q}^{(0)} = \begin{bmatrix} \hat{Q}_1^{(0)} & \hat{Q}_2^{(0)} \end{bmatrix}$ is orthogonal,
3. $\hat{\mathcal{L}}_1 = (\hat{Q}^{(0)})^\top \hat{\mathcal{L}} \hat{Q}^{(0)}$,

4. For $i \in 1:d$,

$$\hat{D}^{(i)} \in \operatorname{ri}\left(\hat{\mathcal{L}}_i^{\perp} \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{r_i} \end{bmatrix}^*\right) = \operatorname{ri}\left(\hat{\mathcal{L}}_i^{\perp} \cap \left\{ X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} : X_{22} \succeq 0 \right\}\right),$$

5. $\hat{D}_{22}^{(d)} = 0$;

6. for $i = 1 : d-1$,

$$\begin{aligned} &\operatorname{range}(\hat{D}_{22}^{(i)}) = \operatorname{range}(\bar{Q}_1^{(i)}), \quad \ker(\hat{D}_{22}^{(i)}) = \operatorname{range}(\bar{Q}_2^{(i)}), \quad \bar{Q}^{(i)} = \begin{bmatrix} \bar{Q}_1^{(i)} & \bar{Q}_2^{(i)} \end{bmatrix}, \\ &\hat{Q}_1^{(i)} = \begin{bmatrix} I_{r-r_i} & 0 \\ 0 & \bar{Q}_1^{(i)} \end{bmatrix}, \quad \hat{Q}_2^{(i)} = \begin{bmatrix} 0 \\ \bar{Q}_2^{(i)} \end{bmatrix}, \quad \hat{Q}^{(i)} = \begin{bmatrix} \hat{Q}_1^{(i)} & \hat{Q}_2^{(i)} \end{bmatrix} \in \mathbb{R}^{r \times r}, \\ &\hat{\mathcal{L}}_{i+1} = (\hat{Q}^{(i)})^\top \hat{\mathcal{L}}_i \hat{Q}^{(i)}. \end{aligned}$$

For $i \in 0 : d-1$, define

$$D^{(i)} := \begin{bmatrix} \hat{D}^{(i)} & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{S}^n, \quad Q_1^{(i)} := \begin{bmatrix} \hat{Q}_1^{(i)} \\ 0 \end{bmatrix}, \quad Q_2^{(i)} := \begin{bmatrix} \hat{Q}_2^{(i)} & 0 \\ 0 & I \end{bmatrix}, \quad Q^{(i)} := \begin{bmatrix} \hat{Q}^{(i)} & 0 \\ 0 & I \end{bmatrix} \in \mathbb{R}^{n \times n},$$

$\bar{\mathcal{L}}^{(1)} := (Q^{(0)})^\top \bar{\mathcal{L}}(Q^{(0)})$ and $\bar{\mathcal{L}}^{(i+1)} := (Q^{(i)})^\top \bar{\mathcal{L}}^{(i)}(Q^{(i)})$ for $i = 1 : d$. Then

- $D^{(0)} \succeq 0$, and $\langle D^{(0)}, C \rangle = \langle \hat{D}^{(0)}, \hat{C} \rangle = 0$ indicates that $D^{(0)} \in \bar{\mathcal{L}}^{\perp}$. (Hence $\bar{\mathcal{L}} \cap \mathbb{S}_{++}^n = \varnothing$.) In fact, if $D \in \bar{\mathcal{L}}^{\perp} \cap \mathbb{S}_+^n$, then $\langle D, V \rangle = (\mathcal{A}(D))^\top v = 0$, so the structure of V given in (7.32) means that $D = \begin{bmatrix} \hat{D} & 0 \\ 0 & 0 \end{bmatrix}$, where $\hat{D} \succeq 0$. Moreover, $\hat{D} \in \hat{\mathcal{L}}^{\perp}$ (because, for instance, $\langle \hat{D}, \hat{C} \rangle = \langle D, C \rangle = 0$). Hence $\bar{\mathcal{L}}^{\perp} \cap \mathbb{S}_+^n = \begin{bmatrix} \hat{\mathcal{L}}^{\perp} \cap \mathbb{S}_+^r & 0 \\ 0 & 0 \end{bmatrix}$. Therefore we have $D^{(0)} \in \operatorname{ri}(\bar{\mathcal{L}}^{\perp} \cap \mathbb{S}_+^n)$.
- It is immediate that $\operatorname{range}(D^{(0)}) = \operatorname{range}(Q_1^{(0)})$ and $\ker(D^{(0)}) = \operatorname{range}(Q_2^{(0)})$.
- For $i \in 1 : d$, we have $D^{(i)} \in \bar{\mathcal{L}}_i^{\perp} \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{n_i} \end{bmatrix}^*$, where $n_i = n - r + r_i$: it is immediate that $D^{(i)} \in \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{n_i} \end{bmatrix}^*$ because $D^{(i)}$ is formed by augmenting zero blocks to $\hat{D}^{(i)} \in \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{r_i} \end{bmatrix}^*$. To

see that $D^{(i)} \in \bar{\mathcal{L}}_i^\perp$, note that

$$\begin{aligned}
\hat{D}^{(i)} \in \hat{\mathcal{L}}_i^\perp &\implies \left(\hat{Q}^{(0)} \cdots \hat{Q}^{(i-1)} \hat{Q}^{(i)}\right) \hat{D}^{(i)} \left(\hat{Q}^{(0)} \cdots \hat{Q}^{(i-1)} \hat{Q}^{(i)}\right)^\top \in \hat{\mathcal{L}}^\perp \\
&\implies \begin{bmatrix} \left(\hat{Q}^{(0)} \cdots \hat{Q}^{(i-1)} \hat{Q}^{(i)}\right) \hat{D}^{(i)} \left(\hat{Q}^{(0)} \cdots \hat{Q}^{(i-1)} \hat{Q}^{(i)}\right)^\top & 0 \\ 0 & 0 \end{bmatrix} \in \bar{\mathcal{L}}^\perp \\
&\implies \left(\prod_{j=0}^{i} \begin{bmatrix} \hat{Q}^{(j)} & 0 \\ 0 & I \end{bmatrix}\right) \begin{bmatrix} \hat{D}^{(i)} & 0 \\ 0 & 0 \end{bmatrix} \left(\prod_{j=0}^{i} \begin{bmatrix} \hat{Q}^{(j)} & 0 \\ 0 & I \end{bmatrix}\right)^\top \in \bar{\mathcal{L}}^\perp \\
&\implies D^{(i)} = \begin{bmatrix} \hat{D}^{(i)} & 0 \\ 0 & 0 \end{bmatrix} \in \left(\prod_{j=0}^{i} \begin{bmatrix} \hat{Q}^{(j)} & 0 \\ 0 & I \end{bmatrix}\right)^\top \bar{\mathcal{L}}^\perp \left(\prod_{j=0}^{i} \begin{bmatrix} \hat{Q}^{(j)} & 0 \\ 0 & I \end{bmatrix}\right) = \bar{\mathcal{L}}_i^\perp.
\end{aligned}$$

In fact, we have $D^{(i)} \in \operatorname{ri}\left(\bar{\mathcal{L}}_i^\perp \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{n_i} \end{bmatrix}^*\right)$, because

$$\begin{aligned}
\left(\prod_{j=0}^{i} \begin{bmatrix} \hat{Q}^{(j)} & 0 \\ 0 & I \end{bmatrix}\right)^\top V \left(\prod_{j=0}^{i} \begin{bmatrix} \hat{Q}^{(j)} & 0 \\ 0 & I \end{bmatrix}\right) &= \left(\prod_{j=0}^{i} \begin{bmatrix} \hat{Q}^{(j)} & 0 \\ 0 & I \end{bmatrix}\right)^\top \begin{bmatrix} 0 & 0 \\ 0 & I_n - r \end{bmatrix} \left(\prod_{j=0}^{i} \begin{bmatrix} \hat{Q}^{(j)} & 0 \\ 0 & I \end{bmatrix}\right) \\
&= \begin{bmatrix} 0 & 0 \\ 0 & I_n - r \end{bmatrix} = V,
\end{aligned}$$

so that $D \in \bar{\mathcal{L}}_i^\perp \cap \begin{bmatrix} 0 & 0 \\ 0 & \mathbb{S}_+^{n_i} \end{bmatrix}^*$ implies that $D_{jk} = 0$ for all $j, k \in (n-r+1) : n$.

The above four points show the first d iterations of facial reduction on $\bar{\mathcal{L}} \cap \mathbb{S}_+^n$. Hence $\mathrm{d}(\bar{\mathcal{L}}) \geq \mathrm{d}(\hat{\mathcal{L}})$. □

Since rotation does not change the degree of singularity (Proposition 7.5.6), we can drop the assumption that $V = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}$ and allow for general V that is singular and nonzero:

Corollary 7.5.8. *Suppose that* (P) *is feasible and there exist* $V \in \mathbb{S}^n$ *and* $v \in \mathbb{R}^m$ *such that*

$$0 \neq V = \mathcal{A}^* v \succeq 0, \quad b^\top v = 0, \tag{7.32}$$

and $\ker(V) = \operatorname{range}(P)$, *where* $P \in \mathbb{R}^{n \times r}$ *has orthonormal columns. Let* $\hat{C} = P^\top C P \in \mathbb{S}^r$, $\hat{A}_i = P^\top A_i P \in \mathbb{S}^r$ *for* $i \in 1 : m$, *and define* $\hat{\mathcal{A}} : \mathbb{S}^r \to \mathbb{R}^m$ *using* $\hat{A}_1, \ldots, \hat{A}_m$. *Define* $\hat{\mathcal{L}} := \operatorname{span}(\hat{C}, \hat{A}_1, \ldots, \hat{A}_m)$. *Then* $\mathrm{d}(\hat{\mathcal{L}}) \leq \mathrm{d}(\bar{\mathcal{L}})$.

Now we can easily prove Lemma 7.5.3.

Proof of Lemma 7.5.3. The minimal face $P\mathbb{S}^r_+P^\top$ can be obtained via facial reduction on (D). At each step of the facial reduction, the new primal (P) remains feasible and the degree of singularity of the linear subspace range$(\mathcal{A}^*_C)$ defining the primal feasible region does not increase, by Corollary 7.5.8. In particular, the projection $P^\top \cdot P$ on the primal feasible region using the minimal face of (D) does not increase the degree of singularity. □

Chapter 8

Classes of problems that fail the Slater condition

As mentioned in Theorem 1.1.1, the Slater condition is a generic property. Nonetheless, in practice the failure of the Slater condition is not as rare as we hope; while we probably do not have to worry too much if the problem data is completely random given the result in Theorem 1.1.1, a number of structured semidefinite programming problems arising from applications are indeed proven to fail the Slater condition. In this chapter, we review some classes of structure SDP that are known to fail the Slater condition and the facial reduction technique can be employed to regularize and reduce the problem size. In addition, knowledge of the minimal face often sheds light on subtle algebraic properties of the SDP.

In this chapter, we give an overview from existing literature of some problems which can be preprocessed by using the facial reduction technique. The problems include:

- SDP relaxation of quadratic assignment problem [105];
- SDP relaxation of traveling salesman problem [29];
- side chain positioning problem [24, 25];
- finding sparse sum-of-squares representation of polynomials [58];
- sensor network localization problem [59, 60];
- Lyapunov equation (see e.g., [21]).

8.1 Symmetric quadratic assignment problem

Suppose we have a list of n facilities that we wish to put in n given distinct locations. The distance between locations i and j is d_{ij} and the flow between facilities i and j is going to be f_{ij}, for all $i, j \in 1:n$. The goal of the quadratic assignment problem (QAP) is to assign each of the n facilities to exactly one of the n given locations in such a way that the total cost of transportation among all the facilities is minimized. In other words, the goal is to find a permutation $\bar{\phi} : (1:n) \to (1:n)$ that solves

$$\min_{\phi \text{ is a permutation of } 1:n} \sum_{i,j=1}^{n} f_{ij} d_{\phi(i),\phi(j)}. \tag{8.1}$$

For simplicity, we omit the linear term in this formulation; see [23] for further details. The combinatorial optimization problem (8.1) can be formulated in terms of permutation matrices. Define $D = [d_{ij}], F = [f_{ij}] \in \mathbb{R}^{n\times n}$; then (8.1) is equivalent to the optimization problem

$$v_{\text{QAP}} := \min_{P\in\Pi^n} \operatorname{tr}(FPD^\top P^\top), \tag{8.2}$$

where Π^n denote the set of all $n \times n$ permutation matrices. We will assume that both F and D are symmetric[1].

An SDP relaxation of (8.2) was proposed in [105]. The idea is to observe that $P \in \Pi^n$ if and only if P satisfies the quadratic equations

$$P \circ P = P, \quad PP^\top = I = P^\top P, \quad \|P\bar{e}_n - \bar{e}_n\|^2 + \|P^\top \bar{e}_n - \bar{e}_n\|^2 = 0, \tag{8.3}$$

where $\bar{e}_n \in \mathbb{R}^n$ is the vector of all ones. (We drop the subscript when the size is clear.) Using

$$PP^\top = \sum_{k=1}^{n} P_{:k} P_{:k}^\top \quad \text{and} \quad [P^\top P]_{ij} = P_{:i}^\top P_{:j}$$

and Kronecker product[2], it is not hard to show that P satisfies (8.3) if and only if $x := \operatorname{vec}(P) = (P_{:,1}; P_{:,2}; \ldots, P_{:,n}) \in \mathbb{R}^{n^2}$ satisfies

$$\operatorname{diag}(xx^\top) = x \circ x = x, \quad \operatorname{bdiag}(xx^\top) = I, \quad \operatorname{odiag}(xx^\top) = I,$$
$$\langle I \otimes \bar{e}\bar{e}^\top + \bar{e}\bar{e}^\top \otimes I, xx^\top\rangle - 4(\bar{e}\otimes\bar{e})^\top x + 2n = 0, \qquad \langle I \otimes (\bar{e}\bar{e}^\top - I) + (\bar{e}\bar{e}^\top - I) \otimes I, xx^\top\rangle = 0,$$

where the last constraint makes use of the fact that $P_{:i}P_{:j}^\top$ is a diagonal matrix if $i = j$ or has a zero diagonal if $i \neq j$; bdiag is a linear map that sums up the diagonal blocks and odiag is a

[1] In fact, assuming only one of F and D to be symmetric suffices for the following discussion.

[2] Recall that for matrices $A = [a_{ij}]_{i\in 1:p, j\in 1:q} \in \mathbb{R}^{p\times q}, B \in \mathbb{R}^{r\times s}$, the *Kronecker product* of A and B is defined as the $pr \times qs$ matrix $[a_{ij}B]_{i\in 1:p,\, j\in 1:q}$.

linear map that returns the trace of each block; explicitly, bdiag and odiag are defined by

$$\mathrm{bdiag} : \mathbb{S}^{n^2} \to \mathbb{S}^n : \begin{array}{c} \\ n \\ n \\ \\ n \end{array} \begin{array}{c} \begin{array}{cccc} n & n & & n \end{array} \\ \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{bmatrix} \end{array} \mapsto \quad \sum_{k=1}^{n} S_{kk},$$

$$\mathrm{odiag} : \mathbb{S}^{n^2} \to \mathbb{S}^n : \begin{array}{c} \\ n \\ n \\ \\ n \end{array} \begin{array}{c} \begin{array}{cccc} n & n & & n \end{array} \\ \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{n1} & S_{n2} & \cdots & S_{nn} \end{bmatrix} \end{array} \mapsto \begin{bmatrix} \mathrm{tr}(S_{11}) & \mathrm{tr}(S_{12}) & \cdots & \mathrm{tr}(S_{1n}) \\ \mathrm{tr}(S_{21}) & \mathrm{tr}(S_{22}) & \cdots & \mathrm{tr}(S_{2n}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{tr}(S_{n1}) & \mathrm{tr}(S_{n2}) & \cdots & \mathrm{tr}(S_{nn}) \end{bmatrix}.$$

(We remark that bdiag and odiag are often called *partial traces*.) Therefore (8.2) is equivalent to

$$\begin{array}{ll} \min\limits_{x,X} & x^\top (D \otimes F) x \\ \text{s.t.} & \mathrm{diag}(X) = x, \\ & \mathrm{bdiag}(X) = I, \quad \mathrm{odiag}(X) = I, \\ & \langle I \otimes \bar{e}\bar{e}^\top + \bar{e}\bar{e}^\top \otimes I, X \rangle - 4(\bar{e} \otimes \bar{e})^\top x + 2n = 0, \\ & \langle I \otimes (\bar{e}\bar{e}^\top - I) + (\bar{e}\bar{e}^\top - I) \otimes I, X \rangle = 0, \\ & X = xx^\top, \quad x \in \mathbb{R}^{n^2}, \ X \in \mathbb{S}^{n^2}, \end{array}$$

which is a rank-constrained SDP (and is NP-hard). Relaxing the rank constraint $X = xx^\top$ to $X \succeq xx^\top$, which is equivalent to $\begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succeq 0$ using Schur complement (Theorem 2.1.5), we obtain an SDP relaxation of the QAP (8.2):

$$\begin{array}{lll} v_{\mathrm{QAP}} \geq v_{\mathrm{QAP-SDP}} & \inf\limits_{x,X,Y} & x^\top (D \otimes F) x \\ & \text{s.t.} & \mathrm{diag}(X) = x, \\ & & \mathrm{bdiag}(X) = I, \quad \mathrm{odiag}(X) = I, \\ & & \left\langle \begin{bmatrix} 2n & -2\bar{e}^\top \otimes \bar{e}^\top \\ -2\bar{e} \otimes \bar{e} & I \otimes \bar{e}\bar{e}^\top + \bar{e}\bar{e}^\top \otimes I \end{bmatrix}, \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \right\rangle = 0, \\ & & \langle I \otimes (\bar{e}\bar{e}^\top - I) + (\bar{e}\bar{e}^\top - I) \otimes I, X \rangle = 0, \\ & & Y = \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succeq 0, \quad x \in \mathbb{R}^{n^2}, \ X \in \mathbb{S}^{n^2}. \end{array} \tag{8.4}$$

Note that the SDP relaxation (8.4) of any instance of the QAP (8.2) of a fixed problem size n

always has the same feasible region. Moreover, the coefficient matrix

$$R := \begin{bmatrix} 2n & -2\bar{e}^\top \otimes \bar{e}^\top \\ -2\bar{e} \otimes \bar{e} & I \otimes \bar{e}\bar{e}^\top + \bar{e}\bar{e}^\top \otimes I \end{bmatrix} = \begin{bmatrix} -\bar{e}_{2n}^\top \\ T^\top \end{bmatrix} \begin{bmatrix} -\bar{e}_{2n} & T \end{bmatrix}, \quad \text{where } T := \begin{bmatrix} I \otimes \bar{e}^\top \\ \bar{e}^\top \otimes I \end{bmatrix} \in \mathbb{R}^{2n \times n^2},$$

i.e., the coefficient matrix R is positive semidefinite, and any (x, X, Y) feasible for (8.4) must satisfy $Y \in \mathbb{S}_+^{n^2} \cap \{R\}^\perp \lhd \mathbb{S}_+^{n^2}$, or equivalently, $Y = W\hat{X}W^\top$ for some $\hat{X} \succeq 0$, where

$$W = \begin{array}{c} \\ 1 \\ n^2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & (n-1)^2 \end{array} \\ \begin{bmatrix} 1 & 0 \\ \frac{1}{n}\bar{e}_n \otimes \bar{e}_n & B_n \otimes B_n \end{bmatrix} \end{array}, \quad \text{where } B_n := \begin{bmatrix} I_{n-1} \\ -\bar{e}^\top \end{bmatrix} \in \mathbb{R}^{n \times (n-1)},$$

is a full column rank matrix satisfying $\mathrm{range}(W) = \ker(R)$ [105, Theorem 3.1]. Using the substitution $Y = W\hat{X}W^\top$, one can reduce the problem size of (8.4), and obtain the smaller equivalent SDP (in variable $\hat{X} \in \mathbb{S}^{(n-1)^2+1}$):

$$\begin{array}{rll} v_{\mathrm{QAP-SDP}} = & \inf\limits_{Y, \hat{X}} & \left\langle \begin{bmatrix} 0 & 0 \\ 0 & D \otimes F \end{bmatrix}, Y \right\rangle \\ & \text{s.t.} & \mathrm{diag}(Y_{2:n,2:n}) = Y_{2:n,1},\ Y_{1,1} = 1, \\ & & \langle I \otimes (\bar{e}\bar{e}^\top - I) + (\bar{e}\bar{e}^\top - I) \otimes I,\ Y_{2:n,\,2:n} \rangle = 0, \\ & & Y = W\hat{X}W^\top,\ \ \hat{X} \in \mathbb{S}^{(n-1)^2+1}. \end{array} \tag{8.5}$$

8.1.1 Slater condition for SDP relaxations of integer programs

We remark that, given a combinatorial optimization problem, the dimension of its feasible region is closely related to the Slater condition for an SDP relaxation of that combinatorial optimization problem.

Very often, the feasible region of a combinatorial optimization problem can be put in the *homogeneous equality form*

$$\hat{\mathcal{P}} := \left\{ x \in \mathbb{R}^n : \hat{\mathcal{A}} \left(\begin{bmatrix} 1 & x^\top \\ x & xx^\top \end{bmatrix} \right) = 0 \right\} \tag{8.6}$$

for some linear map $\hat{\mathcal{A}} : \mathbb{S}^{n+1} \to \mathbb{R}^m$; more generally, any nonempty semialgebraic set, i.e., any set of solutions to a consistent system of finitely many polynomial inequalities, can be posed in the form of (8.6) for some linear map $\tilde{\mathcal{A}}$ [90, Proposition 2.31]. The set (8.6) admits an SDP relaxation in the sense that

$$\hat{\mathcal{P}} \subseteq \mathrm{conv}(\hat{\mathcal{P}}) \subseteq \hat{\mathcal{F}} := \left\{ \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \in \mathbb{S}^{n+1} : \hat{\mathcal{A}} \left(\begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \right) = 0,\ \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succeq 0 \right\}.$$

The Slater condition holds for the SDP relaxation $\hat{\mathcal{F}}$, i.e., the set $\hat{\mathcal{F}}$ contains a positive definite matrix, if and only if $\mathrm{conv}(\hat{\mathcal{P}}) \subseteq \mathbb{R}^n$ is full dimensional [89, Theorem 3.1]. If $\dim(\mathrm{conv}(\hat{\mathcal{P}})) = d < n$, i.e.,

$$x \in \hat{\mathcal{P}} \implies x \in \ell + \mathrm{range}(L^\top),$$

for some $\ell \in \mathbb{R}^n$ and $L \in \mathbb{R}^{d\times n}$, then ℓ and L characterize the minimal face of $\mathbb{S}^n_+$ containing $\hat{\mathcal{F}}$, in the sense that

$$\left\{ \begin{bmatrix} 1 & u^\top \\ u & U \end{bmatrix} \in \mathbb{S}^{d+1} : \mathcal{A}\left(\begin{bmatrix} 1 & 0 \\ \ell & L^\top \end{bmatrix} \begin{bmatrix} 1 & u^\top \\ u & U \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \ell & L^\top \end{bmatrix}^\top \right) = 0,\ \begin{bmatrix} 1 & u^\top \\ u & U \end{bmatrix} \succeq 0 \right\}$$

contains a positive definite matrix [90, Theorem 2.33], implying that

$$\mathrm{face}(\hat{\mathcal{F}},\, \mathbb{S}^{n+1}_+) = \begin{bmatrix} 1 & 0 \\ \ell & L^\top \end{bmatrix} \mathbb{S}^{d+1}_+ \begin{bmatrix} 1 & 0 \\ \ell & L^\top \end{bmatrix}^\top .$$

If the feasible region of a combinatorial optimization problem can be put in the form

$$\mathcal{P} := \left\{ x \in \mathbb{R}^n : \mathcal{A}(xx^\top) = b \right\} \tag{8.7}$$

for some linear map $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ and $b \in \mathbb{R}^m$, then a typical way of obtaining an SDP relaxation of the combinatorial optimization problem is to replace the rank-one positive semidefinite matrix $xx^\top$ by $X \succeq 0$ without rank restriction, so the SDP relaxation would have the feasible region

$$\mathcal{F} := \{ X \in \mathbb{S}^n : \mathcal{A}(X) = b,\ X \succeq 0 \}. \tag{8.8}$$

For $\mathcal{F}$ to contain a positive definite matrix, it is necessary and sufficient that $\mathcal{P}$ contains a basis of $\mathbb{R}^n$:

Theorem 8.1.1. *[89, Theorem 4.1] Consider the sets $\mathcal{P}$ and $\mathcal{F}$ defined in* (8.7) *and* (8.8) *respectively. The Slater condition holds for $\mathcal{F}$ if and only if there exists a linearly independent set $\left\{ v^{(1)}, v^{(2)}, \ldots, v^{(n)} \right\} \subseteq \mathcal{P}$.*

In particular, if there exist a positive semidefinite matrix $A \in \mathbb{S}^n_+$ and a matrix $Q \in \mathbb{R}^{n\times\bar{n}}$ of full column rank such that

$$\mathrm{range}(Q) = \ker(A), \quad \text{and} \quad x^\top A x = 0,\ \forall\, x \in \mathcal{P},$$

then $\mathcal{P} = \left\{ x \in \mathbb{R}^n : \mathcal{A}(xx^\top) = b,\ x^\top A x = 0 \right\}$, and the set

$$\bar{\mathcal{F}} := \{ X \in \mathbb{S}^n : \mathcal{A}(X) = b,\ \langle A, X \rangle = 0,\ X \succeq 0 \} \subseteq \mathcal{F}$$

is a potentially tighter SDP relaxation of $\mathcal{P}$ than $\mathcal{F}$. Note, however, that $\bar{\mathcal{F}} \subseteq \mathbb{S}^n_+ \cap \{A\}^\perp = Q\mathbb{S}^{\bar{n}}_+ Q^\top$, so the tighter SDP relaxation $\bar{\mathcal{F}}$ fails the Slater condition.

Naturally, if one knew of the existence of such a positive semidefinite matrix A, it is possible to formulate an SDP relaxation that does not fail the Slater condition. In fact,

$$\mathcal{P} = \left\{x \in \mathbb{R}^n : \mathcal{A}(xx^\top) = b,\ x \in \mathrm{range}(Q)\right\} = \left\{Qu \in \mathbb{R}^n : \mathcal{A}(Quu^\top Q^\top) = b,\ u \in \mathbb{R}^{\bar{n}}\right\},$$

so yet another SDP relaxation of $\mathcal{P}$ would simply be replacing $uu^\top$ by $U \in \mathbb{R}^{\bar{n}}$:

$$\bar{\mathcal{F}} := \left\{U \in \mathbb{S}^{\bar{n}} : \mathcal{A}(QUQ^\top) = b,\ U \succeq 0\right\}.$$

In this way, the combinatorial optimization problem itself is preprocessed even before forming the SDP relaxation.

8.2 Traveling salesman problem

The famous traveling salesman problem (TSP) is, given n locations and their pairwise distances, to find a shortest path that visits each of the n locations exactly once and that starts and ends at the same location. (Observe that the TSP is a special case of the QAP, taking the matrix F in the objective of (8.2) to be the adjacency matrix C_n of the standard n-cycle, given in (8.11).)

Using the adjacency algebra for cycles, de Klerk *et al.* [29] showed that, given $D \in \mathbb{S}^n$, the SDP

$$\begin{array}{ll} \inf\limits_{X^{(1)},\ldots,X^{(d)}} & \frac{1}{2}\langle D, X^{(1)}\rangle \\ \text{s.t.} & X^{(k)} \geq 0,\ \forall\, k \in 1:d, \\ & \sum_{j=1}^d X^{(j)} = \bar{e}\bar{e}^\top - I, \\ & I + \sum_{j=1}^d \cos\left(\frac{2\pi jk}{n}\right) X^{(j)} \succeq 0,\ \forall\, k \in 1:d, \\ & X^{(k)} \in \mathbb{S}^n,\ \forall\, k \in 1:d, \end{array} \tag{8.9}$$

where $\bar{e}$ is the vector of all ones of appropriate length and $d := \lfloor \frac{n}{2} \rfloor$, provides a lower bound for the optimal value of the symmetric traveling salesman problem:

$$\min_{P \in \Pi^n} \frac{1}{2}\,\mathrm{tr}(DPC_nP^\top), \tag{8.10}$$

where Π^n is the set of all $n \times n$ permutation matrices and $C_n \in \mathbb{S}^n$ is the adjacency matrix of the

standard n-cycle, i.e.,

$$C_n = \begin{bmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}. \tag{8.11}$$

The set $\left\{PC_nP^\top : P \in \Pi^n\right\}$ collects the adjacency matrices of all possible cycles on the complete graph K_n on n vertices. Given any cycle $\mathcal{C}$ on K_n, the matrices $X^{(k)} \in \mathbb{S}^n$ (for each $k \in 1:d$) defined by

$$X^{(k)}_{ij} = \begin{cases} 1 & \text{if the distance between vertices } i \text{ and } j \text{ on } \mathcal{C} \text{ is } k, \\ 0 & \text{otherwise} \end{cases}$$

give a feasible solution $(X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ of (8.9).

We show that the Slater condition fails for (8.9). We find a proper face of $\mathbb{S}^d_+ \times \mathbb{S}^d_+ \times \cdots \times \mathbb{S}^d_+$ containing the feasible slacks of (8.9), and use that proper face to show that the variables $X^{(1)}, X^{(2)}, \ldots, X^{(d)}$ satisfy certain linear equalities (as in (8.12)). First we prove a simple lemma.

Lemma 8.2.1. *If n is odd, then for every $k \in 1:d$,*

$$\sum_{j=1}^{d} \cos\left(\frac{2\pi jk}{n}\right) = -\frac{1}{2}.$$

If n is even, then for every $k \in 1:d$,

$$\sum_{j=1}^{d-1} \cos\left(\frac{2\pi jk}{n}\right) = -\frac{1}{2}\left(1 + (-1)^k\right) = \begin{cases} -1 & \textit{if } k \textit{ is even}, \\ 0 & \textit{if } k \textit{ is odd}. \end{cases}$$

Proof. Note that $\sum_{j=1}^{n} \cos\left(\frac{2\pi jk}{n}\right) = 0$ for all $0 < k < n$: since $\exp\left(\sqrt{-1}\frac{2\pi k}{n}\right) \neq 1$, we have

$$\sum_{j=1}^{n} \exp\left(\sqrt{-1}\frac{2\pi jk}{n}\right) = \exp\left(\sqrt{-1}\frac{2\pi k}{n}\right) \frac{\exp\left(\sqrt{-1}\frac{2\pi nk}{n}\right) - 1}{\exp\left(\sqrt{-1}\frac{2\pi k}{n}\right) - 1} = 0,$$

and $\sum_{j=1}^{n} \cos\left(\frac{2\pi jk}{n}\right) = \operatorname{Re}\left(\sum_{j=1}^{n} \exp\left(\sqrt{-1}\frac{2\pi jk}{n}\right)\right) = 0$.

If n is odd, then $2d + 1 = n$. For all $k \in 1:d$,

$$\begin{aligned} 0 &= \sum_{j=1}^{n} \cos\left(\frac{2\pi jk}{n}\right) \\ &= \sum_{j=1}^{d} \cos\left(\frac{2\pi jk}{n}\right) + \sum_{j=d+1}^{n-1} \cos\left(\frac{2\pi jk}{n}\right) + \cos\left(\frac{2\pi nk}{n}\right); \end{aligned}$$

using the change of variable $l = n - j$, we get

$$\sum_{j=d+1}^{n-1} \cos\left(\frac{2\pi jk}{n}\right) = \sum_{l=1}^{d} \cos\left(\frac{2\pi(n-l)k}{n}\right) = \sum_{l=1}^{d} \cos\left(\frac{2\pi lk}{n}\right),$$

so

$$\sum_{j=1}^{d} \cos\left(\frac{2\pi jk}{n}\right) = -\frac{1}{2}\cos(2\pi n) = -\frac{1}{2}.$$

If n is even, then $2d = n$. For all $k \in 1:d$,

$$\begin{aligned}
0 =& \sum_{j=1}^{n} \cos\left(\frac{2\pi jk}{n}\right) \\
=& \sum_{j=1}^{d-1} \cos\left(\frac{2\pi jk}{n}\right) + \sum_{j=d+1}^{n-1} \cos\left(\frac{2\pi jk}{n}\right) + \cos\left(\frac{2\pi dk}{n}\right) + \cos\left(\frac{2\pi nk}{n}\right) \\
=& \sum_{j=1}^{d-1} \cos\left(\frac{2\pi jk}{n}\right) + \sum_{j=d+1}^{n-1} \cos\left(\frac{2\pi jk}{n}\right) + \left(1 + (-1)^k\right);
\end{aligned}$$

using the change of variable $l = n - j$, we get

$$\sum_{j=d+1}^{n-1} \cos\left(\frac{2\pi jk}{n}\right) = \sum_{l=1}^{d-1} \cos\left(\frac{2\pi(n-l)k}{n}\right) = \sum_{l=1}^{d-1} \cos\left(\frac{2\pi lk}{n}\right),$$

so

$$\sum_{j=1}^{d} \cos\left(\frac{2\pi jk}{n}\right) = -\frac{1}{2}\left(1 + (-1)^k\right).$$

□

Proposition 8.2.2. *Let $(X^{(1)}, X^{(2)}, \ldots, X^{(d)})$ be a feasible solution of (8.9), and define*

$$Z^{(k)} = I + \sum_{j=1}^{d} \cos\left(\frac{2\pi jk}{n}\right) X^{(j)} \succeq 0, \quad \forall\, k \in 1:d.$$

Then $Z^{(k)} \in \mathbb{S}^n_+ \cap \left\{\bar{e}\bar{e}^\top\right\}^\perp$ for all $k \in 1:d$.

Proof. Suppose that n is odd. Summing $Z^{(k)}$ for $k \in 1:d$:

$$0 \preceq \sum_{k=1}^{d} Z^{(k)} = dI - \frac{1}{2}\sum_{j=1}^{d} X^{(j)} = \frac{1}{2}(nI - \bar{e}\bar{e}^\top) \in \mathbb{S}^n_+ \cap \left\{\bar{e}\bar{e}^\top\right\}^\perp,$$

where the second equality uses the fact that $\sum_{j=1}^{d} X^{(j)} = \bar{e}\bar{e}^\top - I$ for any feasible solution $(X^{(1)}, X^{(2)}, \ldots, X^{(d)})$. Since $Z^{(k)} \succeq 0$ for all $k \in 1:d$ and $\mathbb{S}^n_+ \cap \left\{\bar{e}\bar{e}^\top\right\}^\perp \triangleleft \mathbb{S}^n_+$, we have that $Z^{(k)} \in \mathbb{S}^n_+ \cap \left\{\bar{e}\bar{e}^\top\right\}^\perp$ for all $k \in 1:d$.

Suppose that n is even. Then

$$Z^{(d)} = I + \sum_{j=1}^{d} (-1)^j X^{(j)},$$

and

$$\begin{aligned}
\sum_{k=1}^{d-1} Z^{(k)} &= (d-1)I + \sum_{j=1}^{d} \sum_{k=1}^{d-1} \cos\left(\frac{2\pi jk}{n}\right) X^{(j)} \\
&= (d-1)I - \frac{1}{2} \sum_{j=1}^{d} \left(1 + (-1)^j\right) X^{(j)} \\
&= (d-1)I - \frac{1}{2}\left(\bar{e}\bar{e}^\top - I + Z^{(d)} - I\right) \\
&= dI - \frac{1}{2}\bar{e}\bar{e}^\top - \frac{1}{2} Z^{(d)},
\end{aligned}$$

implying

$$\sum_{k=1}^{d-1} Z^{(k)} + \frac{1}{2} Z^{(d)} = \frac{1}{2}(nI - \bar{e}\bar{e}^\top) \in \mathbb{S}^n_+ \cap \left\{\bar{e}\bar{e}^\top\right\}^\perp.$$

Hence $Z^{(k)} \in \mathbb{S}^n_+ \cap \left\{\bar{e}\bar{e}^\top\right\}^\perp$ for all $k \in 1:d$. □

Using the proper face $\mathbb{S}^n_+ \cap \left\{\bar{e}\bar{e}^\top\right\}^\perp$, we can regularize the SDP relaxation (8.9). In addition, Proposition 8.2.2 implies that

$$X^{(j)}\bar{e} = \begin{cases} 2\bar{e} & \text{if } j < \frac{n}{2}, \\ \bar{e} & \text{if } j = \frac{n}{2}, \end{cases} \quad \text{for all } j \in 1:d, \tag{8.12}$$

which is not obvious from the formulation (8.9). To see that (8.12) holds, observe that the $\Omega \in \mathbb{S}^d$ defined by $\Omega_{jk} := \cos\left(\frac{2\pi jk}{n}\right)$ for $j, k \in 1, \ldots, d$,

$$\Omega^{-1}\bar{e} = \begin{cases} -2\bar{e} & \text{if } d \neq \frac{n}{2}, \text{ i.e., if } n \text{ is odd}, \\ \begin{bmatrix} -2\bar{e}_{d-1} \\ -1 \end{bmatrix} & \text{if } d = \frac{n}{2}, \text{ i.e., if } n \text{ is even}. \end{cases}$$

For each $k \in 1:d$, $Z^{(k)}\bar{e} = 0$, implying that

$$\sum_{j=1}^{d} \cos\left(\frac{2\pi jk}{n}\right) x^{(j)} = -\bar{e},$$

where $x^{(j)} := X^{(j)}\bar{e}$ for each $j \in 1:d$. Therefore $\Omega \begin{bmatrix} x^{(1)} & x^{(2)} & \cdots & x^{(d)} \end{bmatrix} = -\bar{e}\bar{e}^\top$, and multiplying Ω^{-1} on both sides,

$$\begin{bmatrix} x^{(1)} & x^{(2)} & \cdots & x^{(d)} \end{bmatrix} = -\Omega^{-1}\bar{e}\bar{e}^\top = \begin{cases} 2\bar{e}\bar{e}^\top & \text{if } n \text{ is odd,} \\ \begin{bmatrix} 2\bar{e} & 2\bar{e} & \cdots & 2\bar{e} & \bar{e} \end{bmatrix} & \text{if } n \text{ is even.} \end{cases}$$

This proves (8.12).

8.3 Side chain positioning problem

The *side chain positioning problem* can be modeled as the combinatorial optimization problem

$$\begin{array}{ll} \min\limits_{x} & x^\top E x \\ \text{s.t.} & \sum\limits_{j} v_j^{(k)} = 1, \ \forall\, k \in 1:p, \\ & x = \begin{bmatrix} (v^{(1)})^\top & (v^{(2)})^\top & \cdots & (v^{(p)})^\top \end{bmatrix}^\top \in \{0,1\}^{n_0}. \end{array} \tag{8.13}$$

(The background of this integer quadratic program is given in Section 9.1.) We mention that any x feasible for (8.13) satisfies $\|x\|^2 = p$, so without loss of generality (by adding to E a sufficiently large multiple of the identity matrix) we may assume that the objective is convex quadratic.

The integer quadratic program (8.13) is NP-hard to solve (even though the linear equality constraints are given by a totally unimodular matrix). In fact, (8.13) can be used to model the maximum k-cut problem, a generalization of max-cut problem; see Section 9.1.2 and Theorem 9.1.2 on Page 133.

Chazelle *et al.* [25] proposed an SDP relaxation for (8.13); see (9.5). Indeed, the facial reduction can be applied on the SDP relaxation; the regularization reduces the runtime significantly and also makes explicit some hidden algebraic properties of the feasible SDP solutions.

We discuss the side chain positioning problem, its SDP relaxation and the regularization via facial reduction in Chapter 9.

8.4 Sparse sum-of-squares representations of polynomials

In this section we give a brief introduction to the background for finding a sparse sum-of-squares (SOS) representation of a given polynomial over $\mathbb{R}^n$.

Let $\mathbb{N}$ denote the set of all positive integers. A *monomial over* $\mathbb{R}^n$ is a function of the form

$$x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n \;\mapsto\; cx^\alpha := cx_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n},$$

where the *coefficient* $0 \neq c \in \mathbb{R}$ and $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{N}^n$ are fixed. The *degree* of the monomial $cx_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$ is the sum of the powers, $\sum_{j=1}^n \alpha_j$. A *polynomial over* $\mathbb{R}^n$ is a function given by a finite sum of monomials over $\mathbb{R}^n$, and its *degree* is defined to be the maximum of the degrees of the constituent monomials. In particular, given a polynomial f of degree r over $\mathbb{R}^n$, we can write

$$f(x) = f(x_1, x_2, \ldots, x_n) = \sum_{\alpha \in \mathbb{N}_r^n} f_\alpha x^\alpha = \sum_{\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n) \in \mathbb{N}_r^n} f_\alpha x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}, \quad \forall\, x \in \mathbb{R}^n.$$

The *support* of the polynomial f is defined as $\mathrm{supp}(f) := \{\alpha \in \mathbb{N}_r^n : f_\alpha \neq 0\}$.

A polynomial f of degree r over $\mathbb{R}^n$ is a *sum-of-squares polynomial*, or simply *SOS polynomial*, if $f = \sum_{j=1}^k (g^{(j)})^2$ for a finite number of polynomials $g^{(1)}, \ldots, g^{(k)}$ over $\mathbb{R}^n$. In particular, a SOS polynomial always has an even degree and is nonnegative. A polynomial f over $\mathbb{R}^n$ is said to be *SOS-representable* over a subset $\mathcal{G} \subseteq \mathbb{N}_r^n$ if $f = \sum_{j=1}^k (g^{(j)})^2$, and $\mathrm{supp}(g^{(j)}) \subseteq \mathcal{G}$ for all $j \in 1:k$; we call the sum $\sum_{j=1}^k (g^{(j)})^2$ an *SOS-representation* of f over $\mathcal{G}$.

Finding an SOS representation of a polynomial f over $\mathcal{G}$ is equivalent to solving the semidefinite feasibility problem [28, 67, 72]:

$$\text{find} \quad V \in \mathbb{S}_+^{|\mathcal{G}|} \quad \text{s.t.} \quad f(x) = \sum_{\beta, \gamma \in \mathcal{G}} V_{\beta,\gamma} x^{\beta+\gamma}. \tag{8.14}$$

(See e.g. [58] for a proof.) By comparing coefficients, it is easy to see that (8.14) is equivalent to the following feasibility problem:

$$\text{find} \quad V \in \mathbb{S}_+^{|\mathcal{G}|} \quad \text{s.t.} \quad f_\alpha = \sum_{\substack{\beta, \gamma \in \mathcal{G}, \\ \beta+\gamma=\alpha}} V_{\beta,\gamma}, \;\forall\, \alpha \in \mathrm{supp}(f). \tag{8.15}$$

Kojima *et al.* [58] and Waki and Muramatsu [94] considered the following problem:

Sparse SOS representation of a polynomial. Suppose a given polynomial f over $\mathbb{R}^n$ is known to be SOS-representable over a subset $\mathcal{G} \subseteq \mathbb{N}_r^n$, but the explicit representation is not given. Determine whether there exists a proper subset $\mathcal{H} \subset \mathcal{G}$ such that f is SOS-representable over $\mathcal{H}$.

The set $\mathcal{G}$ mentioned above can be taken as, e.g. [58, 77],

$$\mathcal{G}^0 := \frac{1}{2} \,\mathrm{conv}\,(\mathrm{supp}(f) \cap (2\mathbb{N}^n)) \cap \mathbb{N}^n.$$

Naturally, if one can find such a subset $\mathcal{H}$, then one can substitute $\mathcal{G}$ with $\mathcal{H}$ in (8.15), resulting in a smaller semidefinite feasibility problem.

[58] showed that if one can find $\mathcal{H}, \mathcal{B}$ such that

$$\varnothing \neq \mathcal{H} \subset \mathcal{G}, \quad \varnothing \neq \mathcal{B} \subset \mathcal{G}, \quad \mathcal{G} = \mathcal{H} \cap \mathcal{B}, \quad (\mathcal{B} + \mathcal{B}) \cap \mathcal{F} = \varnothing, \quad (\mathcal{B} + \mathcal{B}) \cap (\mathcal{G} + \mathcal{H}) = \varnothing, \tag{8.16}$$

then whenever f has an SOS representation over $\mathcal{G}$ given by $\sum_{j=1}^{r} (g^{(j)})^2$ (with $\operatorname{supp}(g^{(j)}) \subseteq \mathcal{G}$ for all $j \in 1:k$), the support of $g^{(j)}$ lies within $\mathcal{H}$ for all $j \in 1:r$, i.e., f is SOS-representable over $\mathcal{H}$. (Indeed, [58, Section 3] showed also that there exists a "smallest" set $\mathcal{G}^*$—with respect to the initial set $\mathcal{G}^0$—over which f has an SOS representation.)

How do we find $\mathcal{G}^*$ or just $\mathcal{H}$ satisfying (8.16) in general? [58] used a graph theoretical approach in finding $\mathcal{G}^*$, though the authors noted that the implementation is not very efficent for large scale problems. [94] observed that the feasibility problem (8.15) fails the Slater condition if there exists a proper subset $\mathcal{H} \subset \mathcal{G}$ such that f is SOS-representable over $\mathcal{H}$. In particular, facial reduction can be applied on (8.15) to find a more sparse SOS representation of f.

8.5 Sensor network localization problem

We consider a "simple" version of the *sensor network localization problem* (ignoring the anchors here):

Sensor network localization problem. Given the number of sensors n and their embedding dimension r, an index set $\mathcal{I} \subseteq \big\{(i,j) \in (1:n)^2 : i < j\big\}$, a collection of known *squared Euclidean distances* $d_{ij} \geq 0$ for $(i,j) \in \mathcal{I}$, find the location of the sensors in $\mathbb{R}^r$ such that their pairwise distances match with the given values, i.e.,

$$\text{find} \quad p^{(1)}, p^{(2)}, \ldots, p^{(n)} \in \mathbb{R}^r \quad \text{s.t.} \quad \|p^{(i)} - p^{(j)}\|^2 = d_{ij}, \ \forall\, (i,j) \in \mathcal{I}. \tag{8.17}$$

Define the linear map

$$\mathcal{K} \ : \ \mathbb{S}^n \to \mathbb{S}^n \ : \ Y \mapsto \operatorname{diag}(Y)\bar{e}^\top + \bar{e}\operatorname{diag}(Y)^\top - 2Y,$$

where $\bar{e} \in \mathbb{R}^n$ is the vector of all ones. It is well-known that the feasibility problem (8.17) can be equivalently phrased as a rank-constrained semidefinite feasibility problem:

Theorem 8.5.1. *The feasibility problem* (8.17) *has a solution if and only if there exists* $X \in \mathbb{S}^n$ *such that*

$$W \circ (\mathcal{K}(X) - D) = 0, \quad X \succeq 0, \quad \operatorname{rank}(X) = r, \tag{8.18}$$

where $W, D \in \mathbb{S}^n$ *are defined by*

$$W_{ij} := \begin{cases} 1 & \text{if } (i,j) \text{ or } (j,i) \in \mathcal{I}, \\ 0 & \text{otherwise;} \end{cases} \qquad D_{ij} := \begin{cases} d_{ij} & \text{if } (i,j) \text{ or } (j,i) \in \mathcal{I}, \\ 0 & \text{otherwise.} \end{cases}$$

Indeed, X *solves* (8.18) *if and only if* $X = PP^\top$ *with* $P = \begin{bmatrix} p^{(1)} & \cdots p^{(n)} \end{bmatrix}^\top \in \mathbb{R}^{n\times r}$ *and* $p^{(1)}, \ldots, p^{(n)}$ *satisfies* $\|p^{(i)} - p^{(j)}\|^2 = d_{ij}$ *for all* $(i,j) \in \mathcal{I}$.

Since the sensor network localization problem (or equivalently, the feasibility problem (8.18)) is NP-hard to solve [51, 52, 80], one common way of tackling (8.18) is to relax the rank constraint and consider the semidefinite feasibility problem

$$W \circ (\mathcal{K}(X) - D) = 0, \quad X \succeq 0. \tag{8.19}$$

It has been shown [59, 60] that (8.19) fails the Slater condition if there exists a subset $\mathcal{J} \subseteq 1:n$ such that

$$i, j \in \mathcal{J},\ i < j \implies (i,j) \in \mathcal{I},$$

i.e., for any $i, j \in \mathcal{J}$, the distance between any two sensors indexed by i and j is known. (The sensors indexed by $\mathcal{J}$ can be thought of as forming a clique.) Using $D_{\mathcal{J},\mathcal{J}}$, it is possible to compute a proper face of $\mathbb{S}^n_+$ containing the feasible solutions of (8.19).

While the minimal face of $\mathbb{S}^n_+$ containing the feasible solutions of (8.18) is dependent on the matrices W and D (meaning that we cannot write down the minimal face as easily as in the case of, e.g., the side chain positioning problem where the feasible region is always the same), it happens that the facial reduction can be implemented without solving conic programs (i.e., the auxiliary problem (5.1)); it is possible to iteratively determine proper faces containing the feasible region of (8.19) using singular value decomposition [59, 60]. This results in an accurate computation of the minimal face of $\mathbb{S}^n_+$ containing the feasible solutions of (8.19).

Facial reduction is especially powerful in preprocessing (8.19), because $D_{\mathcal{J},\mathcal{J}}$, which usually corresponds to pairwise distances in low dimensional space due to the nature of the problem (where the sensors are embedded in low dimensional space), tends to give a low dimensional face of $\mathbb{S}^n_+$ even if the cardinality $|\mathcal{J}|$ is large. Consequently, the problem size collapses rather quickly with the aid of the facial reduction.

8.6 Stability of real square matrices and Lyapunov equation

A matrix $A \in \mathbb{R}^{n\times n}$ is said to be *negative stable* or simply *stable* if all eigenvalues of A have negative real part, and A is said to be *positive stable* if $-A$ is negative stable. Observe that A is

negative stable if and only if its transpose $A^\top$ is negative stable.

It has been shown in [64] that A being negative stable characterizes the asymptotic stability of the linear time-invariant autonomous dynamical system

$$\frac{d}{dt}x(t) = Ax(t),$$

and that the negative stability of A is equivalent to the strict feasibility of a linear matrix inequality:

Theorem 8.6.1. *([55, Theorem 2.2.1]) Let $A \in \mathbb{R}^{n\times n}$.*

(a) A is negative stable if and only if

$$S = A^\top P + PA \prec 0,\ P \succ 0 \tag{8.20}$$

has a solution $(P, S) \in \mathbb{S}^n \times \mathbb{S}^n$.

(b) Suppose that $(P, S) \in \mathbb{S}^n \times \mathbb{S}^n$ satisfies the Lyapunov equation

$$S = A^\top P + PA \tag{8.21}$$

and $S \prec 0$. Then A is negative stable if and only if $P \succ 0$.

□

Interestingly, if A is negative stable, then for each $S \in \mathbb{S}^n$, there is a unique solution to (8.20):

Theorem 8.6.2. *([55, Thoerem 2.2.3]) Let A be negative stable. For each $S \in \mathbb{R}^{n\times n}$,* (8.21) *has a unique solution P. If S is symmetric, then the solution P must be symmetric. If S is negative definite, then P must be positive definite.*

How do we find a solution of (8.20)? It is possible to use linear algebra techniques. (See e.g., [53, Chap. 15].) Alternatively, a "textbook" approach is to set up an appropriate SDP to find a solution of (8.20). Consider the SDP (see, e.g., equation (2.20) in [21, Section 2.2.4]):

$$\begin{array}{rll} v_{\mathrm{P}}^L := & \min\limits_{P,\lambda} & \lambda \\ & \text{s.t.} & A^\top P + PA \preceq \lambda I, \\ & & P \succeq I, \\ & & \lambda \geq -1. \end{array} \tag{8.22}$$

We first show that the SDP (8.22) does indeed help us determine whether (8.20) is solvable.

Proposition 8.6.3. *For any $A \in \mathbb{R}^{n\times n}$, (8.20) has a solution if and only if $v_{\mathrm{P}}^L < 0$.*

Proof. First we point out that (8.20) is always feasible: for any $A \in \mathbb{R}^{n\times n}$, let $P = 2I$ and $\lambda = \max\left\{-1, \lambda_{\max}(A^\top P + PA)\right\} + 1$. Then $P \succ I$, $\lambda > -1$ and $\lambda I \succ \left(\lambda_{\max}(A^\top P + PA)\right)(I) \succeq A^\top P + PA$. Hence (P, λ) is a Slater point of (8.20), and (8.20) is feasible.

If (8.20) has no solution, then for any feasible solution (P, λ) of (8.22), P being positive definite implies that $A^\top P + PA$ is not negative definite, i.e., λ cannot be negative. Hence $v_{\mathrm{P}}^L \geq 0$.

Suppose that (8.20) has a solution (P, S). Since (8.20) is homogeneous, without loss of generality we assume that $P \succeq 0$. Let $\lambda = \max\{\lambda_{\max}(S), -1\} < 0$. Then $\lambda \geq -1$ and $S \preceq \lambda_{\max}(S)I \preceq \lambda I$. Hence (P, λ) is a solution of (8.20), and $v_{\mathrm{P}}^L \leq \lambda < 0$. □

For any $A \in \mathbb{R}^{n\times n}$, $(2I, \max\left\{-1, \lambda_{\max}(A^\top P + PA)\right\}+1)$ is a Slater point of (8.22). Moreover, the constraint $\lambda \geq -1$ in (8.22) guarantees that $v_{\mathrm{P}}^L \geq -1$. Hence strong duality holds for (8.22), i.e., (8.22) and its dual[3]

$$\begin{array}{rll} v_{\mathrm{D}}^L := & \max\limits_{Y,Z,y} & \langle I, Z\rangle - y \\ & \text{s.t.} & AY + YA^\top - Z = 0, \\ & & \langle I, Y\rangle + y = 1 \\ & & Y \succeq 0,\ Z \succeq 0,\ y \geq 0 \end{array} \tag{8.23}$$

have the same optimal value $v_{\mathrm{P}}^L = v_{\mathrm{D}}^L$, and v_{D}^L is attained.

An interesting fact is that if A is negative stable, then (8.23) fails the Slater condition:

Proposition 8.6.4. *The SDP (8.23) satisfies the Slater condition if and only if A is positive stable.*

Proof. First note that A is positive stable if and only if $A^\top$ is also positive stable.

Suppose that A is positive stable. Then by Proposition 8.6.1 there exist $Y, Z \succ 0$ such that $Z = AY + YA^\top$. By scaling Y, Z we may assume that $\mathrm{tr}(Y) = 0.5$. Then $(Y, Z, 0.5)$ is a Slater point of (8.23).

[3] Using

$$\langle Y, A^\top P\rangle = \mathrm{tr}((AY)^\top P) = \langle AY, P\rangle \quad \text{and} \quad \langle Y, PA\rangle = \mathrm{tr}(AYP) = \langle YA^\top, P\rangle,$$

the Lagrangian of (8.22) is given by

$$\begin{aligned} L(P, \lambda, Y, Z, y) =& \lambda + \langle Y, A^\top P + PA - \lambda I\rangle + \langle Z, I - P\rangle - y(1 + \lambda) \\ =& \langle AY + YA^\top - Z, Y\rangle + \lambda(1 - \langle I, Y\rangle - y) + \langle I, Z\rangle - y. \end{aligned}$$

Hence for any Y, Z, y, $\inf_{P,\lambda} L(P, \lambda, Y, Z, y)$ is finite and equals $\langle I, Z\rangle - y$ if and only if $AY + YA^\top - Z = 0$ and $1 - \langle I, Y\rangle - y = 0$. Therefore the dual of (8.22) is given by (8.23).

Conversely, suppose that A is not positive stable. Fix any feasible point (Y, Z, y) of (8.23). If $Z \succ 0$, then Y cannot be positive definite by Proposition 8.6.1. Hence (8.23) has no Slater point. □

If A is negative stable, then A cannot be positive stable and by Proposition 8.6.4, the Slater condition does not hold for (8.23). By Theorem 3.3.10, we know that the set of optimal solutions for (8.22) is either empty or unbounded.

We perform a simple numerical experiment on solving (8.22). We fix the maximum real part of the eigenvalues of $A \in \mathbb{R}^{10\times 10}$ at a specific (negative) value, and solve (8.22) using CVX [48, 56] with SDPT3 [92]. Each column of Table 8.1 contains the average over 20 instances. The row "# success" records the number of instance that SDPT3 can solve without terminating prematurely due to numerical errors., and the row "# iter" records the number of iterations SDPT3 takes to solve (8.22).

Table 8.1: Numerics from randomly generated instances of negative stable matrix A

$\max(\mathrm{Re}(A))$	-1	-0.1	-0.01	-0.001	-0.0001	-0.00001	-0.000001
# success	20	20	20	20	20	19	17
cpu time	0.5425	0.5065	0.5435	0.67	0.7585	1.0045	0.8245
# iter	9	9.45	10.5	14.15	16.65	23.55	18.4
$\|P\|_F$	40.387	208.29	295.27	4902.9	23108	3.4508e+05	1.238e+06
$\lambda_{\max}(P)$	30.663	190.6	266.56	4691.2	22221	3.3658e+05	1.1835e+06
$\lambda_{\min}(P)$	3.3299	4.9007	8.2561	28.768	72.914	954.01	6705.9
$\|S\|_F$	134.4	212.58	313.13	1126.4	3022.6	50689	2.4262e+05
$\lambda_{\max}(S)$	-23.017	-5.9203	-1.3817	-1.6324	-1.0455	-1.0759	-0.77924
$\lambda_{\min}(S)$	-48.335	-85.958	-123.29	-472.58	-1631	-31664	-1.2332e+05

Observe that the norm of the of the solution (P, S) and the number of iterations both go up as $\max(\mathrm{Re}(A))$ approaches zero. We see that formulating an SDP without care (e.g. checking whether the Slater condition fails) can indeed lead to numerical difficulties (in this case the solution norm blowing up).

8.7 Summary

In all the above examples, the facial reduction is used for regularizing SDPs with hidden algebraic properties (i.e., the feasible solutions tend to satisfy some overlooked equations). If a problem is

modeled properly, in the sense that the hidden algebraic properties are accounted for (as in e.g., Section 8.1.1), then the resultant SDP would have probably satisfied the Slater condition.

Chapter 9

Side chain positioning problem

In this chapter, we study the side chain positioning problem (stated in (IQP_{SCP}) below), a combinatorial optimization problem arising from protein folding. The side chain positioning problem has been shown to be NP-hard [1], and Chazelle *et al.* [25] studied an SDP relaxation for solving the combinatorial optimization problem. The SDP relaxation, however, fails the Slater condition, and it is observed empirically that the solution of the SDP relaxation is much slower and less accurate, leading to absurd integral solutions after the rounding.

There are three main goals in this chapter. First we formulate an SDP relaxation of the side chain positioning problem that is at least as strong as the SDP relaxation from [25]. Second, we show that it is possible to regularize the SDP relaxation from [25] (as well as the SDP relaxation that we obtain) using facial reduction, arriving at a smaller and more stable SDP. Finally, we use a cutting plane technique to improve the solution quality and to avoid handling a formiddable amount of inequality constraints, thus ensuring quality and efficiency at the same time.

The organization of this chapter is as follows. In Section 9.1, we introduce the side chain positioning problem. We also point out that the side chain positioning problem encompasses the max k-cut problem, which in turns models a wide range of optimization problems (see, for instance, Ghaddar's master thesis [44]), as a special case. In Section 9.2, we derive an SDP relaxation of the side chain positioning problem. In Section 9.3, we show that the SDP relaxation can be regularized by restricting the feasible solutions on the minimal face of the PSD cone containing them. We will also show that our SDP relaxation is tighter than that proposed in [25]. In Section 9.4, we consider the implementation issues of solving the SDP relaxation, including a review of possible rounding techniques and the cutting plane technique. Numerical tests are presented in Section 9.5; we will also study the biological relevance of our numerical results.

9.1 Introduction to the side chain positioning problem and its connection to max k-cut problem

Consider the integer quadratic programming problem

$$\begin{aligned} v_{\text{scp}} = \quad & \min_{x} \quad x^\top E x \\ & \text{s.t.} \quad \sum_{j \in \mathcal{V}_k} x_j = 1, \quad \forall\, k \in 1:p, \\ & \qquad x \in \{0,1\}^{n_0}, \end{aligned} \tag{IQP$_{\text{SCP}}$}$$

where

- $p, m_1, m_2, \ldots, m_p$ are positive integers;
- $n_0 := \sum_{k=1}^{p} m_k$;
- $E \in \mathbb{S}^{n_0}$
- $\bar{m}_0 := 0$ and $\bar{m}_k := \sum_{l=0}^{k} m_l$ for $k \in 1:p$; and
- $\mathcal{V}_k = (\bar{m}_{k-1} + 1) : \bar{m}_k$.

The integer quadratic program (IQP$_{\text{SCP}}$) models the following combinatorial optimization problem.

Side chain positioning problem. Given an undirected graph with loops and no parallel edges, where the vertex set is given by $\mathcal{V} := \bigcup_{k=1}^{p} \mathcal{V}_k$, the edge set is given by $\mathcal{E}$ and edge weights ω_{ij} for all $\{i,j\} \in \mathcal{E}$, pick exactly one vertex from each partition $\mathcal{V}_k$ (for $k \in 1:p$) such that the sum of edge weights of the subgraph induced by these chosen vertices is minimized.[1]

In this section we give some background information about the combinatorial optimization problem (IQP$_{\text{SCP}}$). We first outline the motivation of (IQP$_{\text{SCP}}$) from the protein folding problem in Section 9.1.1. Then we show that the max k-cut problem, a generalization of the max-cut problem, can in fact be formulated as a special case of (IQP$_{\text{SCP}}$). In particular, many combinatorial optimization problems that can be modeled as a max k-cut problem, such as the Potts glass problem (see [44] for a detailed list of applications), can be formulated in the form of (IQP$_{\text{SCP}}$).

[1] A word on the objective $x^\top E x$: given the edge weights ω_{ij} for all $\{i,j\} \in \mathcal{E}$, define the matrix $E \in \mathbb{S}^{n_0}$ by

$$E_{ij} := \begin{cases} \frac{1}{2}\omega_{ij} & \text{if } \{i,j\} \in \mathcal{E},\ i \neq j, \\ \omega_{ij} & \text{if } \{i,j\} \in \mathcal{E},\ i = j, \\ 0 & \text{if } \{i,j\} \notin \mathcal{E}. \end{cases}$$

9.1.1 Protein folding: the biology behind the side chain positioning problem

The side chain positioning problem is a discretized subproblem of the protein folding problem, which is the subject of this section. (We emphasize that the coverage of the biological background in this chapter is far from complete. We omit, for instance, the issue of problem data generation and the domain knowledge on protein structure used for preprocessing the problem data, e.g., the dead-end elimination. More details on the biology can be found in, e.g., [24, 25].)

We first discuss the basic structure of proteins. Amino acids are the building blocks of a protein; an amino acid is a molecule consisting of an alpha-carbon (-C_α-) acting as the "hub" connecting four atom groups:

(1) a hydrogen atom,

(2) an amino group (-NH_2),

(3) a carboxyl group (-COOH) and

(4) an atom group called the *side chain* (which we represent by -R).

(In other words, any two amino acids only differ in the composition of the side chain.) A protein is a chain of amino acids linked through a condensation reaction: the carboxyl group of an amino acid is linked to the amino group of the next amino acid, and in the process the carboxyl group -COOH gives up -OH while the amino group -NH_2 gives up a hydrogen atom, to produce a water molecule. The result of this condensation reaction is a chain $\cdots \mathrm{NC_\alpha C\ NC_\alpha C\ NC_\alpha C}\cdots$ of repetitive triple atoms $\mathrm{NC_\alpha C}$, linked by CN bonding formed in the condensation reaction, and from each C_α atom a side chain molecule sprouts. We call the chain $\cdots \mathrm{NC_\alpha C\ NC_\alpha C\ NC_\alpha C}\cdots$ the *backbone* of the protein, and each of the repeating units (including the side chains) a *residue* of the protein.

With the basic understanding of protein structure, we can outline the *protein folding problem*. Given the chemical content of a protein, we are often interested in determining the protein conformation, i.e., the three-dimensional positioning of the constituent residues (i.e., the amino acids) of the protein, that is optimal in some way. A protein in its natural form is believed to arrange its constituent residues in a way that minimizes the total energy: between any two residues there is an interaction energy, and each residue itself carries some level of "self-energy" which may change depending on its surroundings. A protein conformation is considered "optimal" if the total energy is minimized. Therefore, the protein folding problem is, given the constituent molecules, to find three dimensional positions of its molecules that minimizes its total energy.

One subproblem of the protein folding problem is the determination of the side chain position: assuming that the positions of atoms in the backbone are known, determine the position of the side chains. This problem is often further simplified by assuming that each of the side chains can take only one of *finitely many* possible positions; we call these possible positions the *rotamers*. With all these assumptions, we arrive at the *side chain positioning problem*: for each residue of the protein, pick exactly one rotamer (positioning of the side chain atoms) so that the total energy of the protein is minimized.

9.1.2 Complexity and relation to max k-cut problem

It has been shown [1] that the integer quadratic program ($\mathrm{IQP}_{\mathrm{SCP}}$) is NP-hard. In fact, it is not even "easy" to find a "good" approximation in the following sense:

Theorem 9.1.1. *[25, Theorem 5.1] It is NP-complete to approximate the optimal value v_{scp} of ($\mathrm{IQP}_{\mathrm{SCP}}$) within a factor of γn_0, where γ is a positive constant (and n_0 is the total number of rotamers).*

In this section, we offer an alternative proof of the known NP-hardness result of the side chain positioning problem [1, 25], by showing that the maximum k-cut (or max k-cut in short) problem can be reduced to ($\mathrm{IQP}_{\mathrm{SCP}}$). In particular, the NP-hardness and inapproximability results of the max-k cut problem apply to ($\mathrm{IQP}_{\mathrm{SCP}}$) too. Recall the max k-cut problem:

Max k-cut problem. Given an undirected graph with vertex set $\tilde{\mathcal{V}}$, edge set $\tilde{\mathcal{E}}$ and edge weights ω_{ij} for all $(i,j) \in \tilde{\mathcal{E}}$, partition $\tilde{\mathcal{V}}$ into k sets so that the total weight of edges with ends on two different partitions is maximized.

Naturally, in the context of the max k-cut problem, we may assume that the graph of interest has no parallel edges (though it may contain loops). As noted in [44], the max k-cut problem is the same as the minimum k-partition (or min k-partition for short) problem:

Min k-partition problem. Given an undirected graph with vertex set $\tilde{\mathcal{V}}$, edge set $\tilde{\mathcal{E}}$ and edge weights ω_{ij} for all $(i,j) \in \tilde{\mathcal{E}}$, partition $\tilde{\mathcal{V}}$ into k sets so that the total weight of edges that have both ends in the same partition is minimized.

(To see that min k-partition problem and max k-cut problem are really the same, simply note that for any partition of $\tilde{\mathcal{V}}$ into k sets, the total weight of edges with ends on two different partitions and the total weight of edges that have both ends in the same partition always sum to the constant $\sum_{(i,j)\in\tilde{\mathcal{E}}} \omega_{ij}$.)

We prove that the min k-partition problem is polynomial-time reducible to the side chain positioning problem.

Theorem 9.1.2. *For any positive integer k, the min k-partition problem is polynomial-time reducible to the side chain positioning problem.*

Proof. Let $\tilde{\mathcal{G}} = (\tilde{\mathcal{V}}, \tilde{\mathcal{E}})$ be an undirected graph with edge weights ω_{v_i,v_j} $(\forall\, \{v_i, v_j\} \in \tilde{\mathcal{E}})$. We construct an undirected weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $\mathcal{V}$ being a disjoint union of $\mathcal{V}_l$ with cardinality k for $l \in 1 : |\tilde{\mathcal{V}}|$ and $|\mathcal{E}| = k|\tilde{\mathcal{E}}|$ so that for any $\alpha \in \mathbb{R}$, the following are equivalent:

(1) $\tilde{\mathcal{V}}$ can be partitioned into k sets with the total weights of the edges with both ends in the same partition no larger than α.

(2) There exists a vertex $v_l \in \mathcal{V}_l$ for each $l \in 1 : k$ such that the total edge weights of the subgraph induced by $v_1, v_2, \ldots, v_k$ is no larger than α.

In fact, writing $\tilde{\mathcal{V}} := \{v_1, v_2, \ldots, v_q\}$, define $\mathcal{V}_j := \left\{v_j^{(1)}, v_j^{(2)}, \ldots, v_j^{(k)}\right\}$ for $j \in 1 : q$. (In other words, $\mathcal{V}_j$ stores k identical copies of vertex v_j.) Define the edge set of $\mathcal{E}$ via

$$\left\{v_i^{(l_1)}, v_j^{(l_2)}\right\} \in \mathcal{E} \iff l_1 = l_2,\ \{v_i, v_j\} \in \tilde{\mathcal{E}}, \quad (\forall\, l_1, l_2 \in 1 : k,\ i, j \in 1 : q), \tag{9.1}$$

and define the weight of any arbitrary edge $\left\{v_i^{(l_1)}, v_j^{(l_2)}\right\} \in \mathcal{E}$ to be ω_{ij}.

Note that the construction of the weighted graph $\mathcal{G}$ involves no computation, so the construction can trivially be done in polynomial time. (Note also that the problem size due to the graph $\mathcal{G}$ is k times the problem size due to the graph $\tilde{\mathcal{G}}$.)

Suppose that $\tilde{\mathcal{V}}$ can be partitioned into $\tilde{\mathcal{V}}_1, \tilde{\mathcal{V}}_2, \ldots, \tilde{\mathcal{V}}_k$ such that

$$\sum_{l=1}^{k} \sum_{\substack{v_i, v_j \in \tilde{\mathcal{V}}_l \\ \{v_i, v_j\} \in \tilde{\mathcal{E}}}} \omega_{ij}. \tag{9.2}$$

Now consider our constructed graph $\mathcal{G}$. For each partition $\mathcal{V}_j$, pick $v_j^{(l_j)}$, where l_j indexes the partition that v_j belongs to in $\tilde{\mathcal{G}}$, i.e., $v_j \in \tilde{\mathcal{V}}_{l_j}$. Therefore

$$v_i^{(l_i)}, v_j^{(l_j)} \text{ satisfy } l_i = l_j \iff v_i, v_j \text{ are in the same partition } \mathcal{V}_l \text{ with } l = l_i = l_j, \tag{9.3}$$

and $(v_1^{(l_1)}, v_2^{(l_2)}, \ldots, v_q^{(l_q)})$ is a feasible solution of the side chain positioning problem; due to the destination of edges in $\mathcal{G}$, its objective value is

$$\sum_{l=1}^{k} \sum_{\substack{l_i = l = l_j, \\ \{v_i, v_j\} \in \tilde{\mathcal{E}}}} \omega_{v_i^{(l_i)}, v_j^{(l_j)}} \tag{9.4}$$

By (9.1) and (9.3), the quantities in (9.2) and (9.4) are the same.

Conversely, suppose that $(v_1^{(l_1)}, v_2^{(l_2)}, \ldots, v_q^{(l_q)})$ is a feasible solution of the side chain positioning problem instance on our constructed graph $\mathcal{G}$. We already saw that its objective value is given by (9.4). For each $l \in 1:k$, let $\tilde{\mathcal{V}}_l := \{v_i : l = l_i\}$. Naturally, the sets $\tilde{\mathcal{V}}_l$ (for $l \in 1:k$) are disjoint. To see that $\tilde{\mathcal{V}} = \bigcup_{l=1}^k \mathcal{V}_l$, simply note that for each $i \in 1:q$, exactly one of element of $\mathcal{V}_i$ is selected. Therefore $\tilde{\mathcal{V}}_1, \tilde{\mathcal{V}}_2, \ldots, \tilde{\mathcal{V}}_k$ is a feasible solution for the min k-partition problem instance on $\tilde{\mathcal{G}}$, and its objective is given in (9.2). Again, the quantities in (9.2) and (9.4) are equal, since $v_i, v_j \in \tilde{\mathcal{V}}_l$ if and only if $v_i^{(l)}, v_j^{(l)}$ are part of the feasible solution of the side chain positioning problem instance.

Therefore statements (1) and (2) at the beginning of the proof are equivalent, and shows that our construction provides a valid reduction of the min k-partition problem into the side chain positioning problem. □

Theorem 9.1.2 highlights the versatility of the side chain positioning problem. Nonetheless, in the remainder of this chapter (in particular in the numerics), we focus on the side chain positioning problem as a protein folding subproblem.

Notation

For any matrices (or vector) S, T of the same size, the *Hadamard product* is defined as a matrix (or vector) of the same size resulting from the element-wise product of the two matrices, denoted by $S \circ T$.

We use $\bar{e}_k$ to denote the vector of all ones in $\mathbb{R}^k$, and $\bar{E}_k$ to denote the $k \times k$ matrix of all ones. When the dimensions are clear, we would omit the subscript k.

9.2 An SDP relaxation of the side chain positioning problem

We already saw in the last section that (IQP_{SCP}) is NP-hard. One common heuristic for approximating NP-hard integer programs is to use an SDP relaxation, which is the subject of this section.

We first list some valid constraints on the variable x in (IQP_{SCP}). Then we derive an SDP relaxation of (IQP_{SCP}). The procedure for deriving the SDP relaxation presented in Section 9.2.2

is similar to that for obtaining the SDP relaxation of ($\mathrm{IQP}_{\mathrm{SCP}}$) in [25], given by:

$$\begin{array}{rll} v_{\mathrm{scp}} \geq & \inf\limits_{Y \in \mathbb{S}^n} & \left\langle \begin{bmatrix} 0 & 0 \\ 0 & E \end{bmatrix}, Y \right\rangle \\ & \text{s.t.} & \sum_{j \in \mathcal{V}_k} Y_{j+1,\, j+1} = 1, \ \forall\, k \in 1:p, \\ & & \sum_{i,j \in \mathcal{V}_k} Y_{i+1,\, j+1} = 1, \ \forall\, k \in 1:p, \\ & & Y_{j+1,\, j+1} - Y_{j+1,\, 1} = 0, \ \forall\, j \in 1:n_0, \\ & & Y_{11} = 1, \\ & & Y \succeq 0, \\ & & Y \geq 0. \end{array} \tag{9.5}$$

Our SDP relaxation uses more valid equality constraints than in (9.5) and fewer nonnegativity constraints on Y. Our SDP relaxation attempts to balance the conflicting goals of having a tight SDP relaxation (by keeping all the constraints $Y_{ij} \geq 0$) and of obtaining an SDP solution in a reasonable amount of time. (See Sections 9.3.3 and the numerics in Section 9.5 for comparison between the two SDP relaxations.)

9.2.1 Valid constraints for the side chain positioning problem

Given any integral vector $\tilde{m} = (\tilde{m}_1, \ldots, \tilde{m}_{\tilde{p}}) \geq 0$, define the matrix

$$A^{\tilde{m}} := \begin{array}{c} \\ 1 \\ 1 \\ \\ 1 \end{array} \begin{array}{c} \begin{array}{cccc} \tilde{m}_1 & \tilde{m}_2 & & \tilde{m}_{\tilde{p}} \end{array} \\ \begin{bmatrix} \bar{e}^\top & 0 & \cdots & 0 \\ 0 & \bar{e}^\top & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{e}^\top \end{bmatrix} \end{array} \in \{0,1\}^{p \times \sum_k \tilde{m}_k}. \tag{9.6}$$

(We take the convention that any matrix $B \in \mathbb{R}^{s \times t}$ is vacuous if s or $t = 0$.) The matrix $A^{\tilde{m}}$ satisfies $(A^{\tilde{m}})^\top \bar{e} = \bar{e} \in \mathbb{R}^{\sum_k \tilde{m}_k}$,

$$(A^{\tilde{m}})^\top A^{\tilde{m}} = \begin{array}{c} \\ \tilde{m}_1 \\ \tilde{m}_2 \\ \\ \tilde{m}_p \end{array} \begin{array}{c} \begin{array}{cccc} \tilde{m}_1 & \tilde{m}_2 & & \tilde{m}_{\tilde{p}} \end{array} \\ \begin{bmatrix} \bar{E} & 0 & \cdots & 0 \\ 0 & \bar{E} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \bar{E} \end{bmatrix} \end{array} \in \mathbb{S}^{\sum_k \tilde{m}_k}, \quad A^{\tilde{m}}(A^{\tilde{m}})^\top = \begin{bmatrix} \tilde{m}_1 & 0 & \cdots & 0 \\ 0 & \tilde{m}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \tilde{m}_{\tilde{p}} \end{bmatrix} \in \mathbb{S}^{\tilde{p}}.$$

When $\tilde{m}$ is taken to be the default input m for ($\mathrm{IQP}_{\mathrm{SCP}}$), we drop the subscript: $A := A^m$. Using A, we can rewrite ($\mathrm{IQP}_{\mathrm{SCP}}$) as

$$\begin{array}{rll} v_{\mathrm{scp}} = & \min\limits_{x} & x^\top E x \\ & \text{s.t.} & Ax - \bar{e} = 0 \in \mathbb{R}^p, \\ & & x \in \{0,1\}^{n_0}, \end{array} \tag{9.7}$$

Observe that if $x \in \{0,1\}^{n_0}$ is feasible for (9.7), then x and the rank one matrix $xx^\top$ satisfy the following constraints:

- $\bar{e}^\top x = p$;
- $\|Ax - \bar{e}\|^2 = 0$, i.e.,
$$x^\top A^\top A x - 2\bar{e}^\top x + p = 0.$$
- Using the simple fact that 0,1 are the solutions of the quadratic equation $t^2 - t = 0$, we get
$$\mathrm{diag}(xx^\top) = x \circ x = x;$$
in fact, the constraint $\mathrm{diag}(xx^\top) = x$ is equivalent to $x \in \{0,1\}^{n_0}$;
- $(A^\top A - I) \circ (xx^\top) = 0$: to see that the $xx^\top$ satisfies the constraint, write
$$x = \left[(v^{(1)})^\top \quad (v^{(2)})^\top \quad \cdots \quad (v^{(p)})^\top \right]^\top, \quad v^{(k)} \in \{0,1\}^{m_k}.$$
x is feasible for ($\mathrm{IQP}_{\mathrm{SCP}}$) if and only if each $v^{(k)}$ has exactly one nonzero entry (which is positive and equals 1). Therefore $v^{(k)}(v^{(k)})^\top$ is a diagonal matrix, i.e., $(\bar{E} - I) \circ (v^{(k)}(v^{(k)})^\top)$ for $k \in 1:p$. This together with $\mathrm{diag}(xx^\top) = x \geq 0$ indicates that

the diagonal blocks of $xx^\top$ are diagonal matrices with nonnegative entries.

In particular,
$$(xx^\top)_{ij} \geq 0, \ \forall\, (i,j) \in \mathcal{B},$$
where
$$\begin{array}{rl} \mathcal{B} := & \left\{ (i,j) : 1 \leq i < j \leq n_0, \ (A^\top A)_{ij} = 1 \right\} \\ = & \left\{ (i,j) : 1 \leq i < j \leq n_0, \ i,j \in \mathcal{V}_k \text{ for some } k \in 1:p \right\}. \end{array} \tag{9.8}$$
- $xx^\top \geq 0$; in particular,
$$(xx^\top)_{ij} \geq 0, \ \forall\, (i,j) \in \mathcal{I}, \tag{9.9}$$

for any index set $\mathcal{I} \subseteq \{(i,j) : 1 \le i < j \le n_0\}$. The set $\mathcal{I}$ will be used to index the cutting planes. As pointed out, the constraints $(A^\top A - I) \circ (xx^\top) = 0$ and $x \circ x = x$ ensure that the diagonal blocks of $xx^\top$ are nonnegative diagonal matrices, we will only consider the nonnegative inequalities $(xx^\top)_{ij} \ge 0$ that occur on the off-diagonal blocks. In other words, we will be interested in the index set $\mathcal{I}$ being a subset of

$$\mathcal{I}_{\ge 0} := \{(i,j) : 1 \le i < j \le n_0,\ (i,j) \notin \mathcal{B},\ i,j \text{ integer}\}, \tag{9.10}$$

where $\mathcal{B}$ defined in (9.8) indexes the diagonal blocks.

For any $\mathcal{I} \subseteq \mathcal{I}_{\ge 0}$, define the projection

$$\bar{\mathcal{P}}_{\mathcal{I}} : \mathbb{S}^{n_0}_+ \to \mathbb{R}^{|\mathcal{I}|} : X \mapsto (X_{ij})_{(i,j)\in\mathcal{I}}. \tag{9.11}$$

Then (9.9) holds if and only if $\bar{\mathcal{P}}_{\mathcal{I}}(xx^\top) \ge 0$. The adjoint $\bar{\mathcal{P}}^*_{\mathcal{I}} : \mathbb{R}^{|\mathcal{I}|} \to \mathbb{S}^{n_0}$ is given as follows: for any $x \in \mathbb{R}^{|\mathcal{I}|}$, $X := \bar{\mathcal{P}}^*_{\mathcal{I}}(x) \in \mathbb{S}^{n_0}$ satisfies $X_{ij} = X_{ji} = \frac{1}{2}x_{ij}$ for all $(i,j) \in \mathcal{I}$, and $X_{ij} = 0$ for all $(i,j) \notin \mathcal{I}$.

Using the valid constraints, we get that (IQP$_{\text{SCP}}$) is equivalent to

$$\begin{array}{rll} v_{\text{scp}} = v_{\mathcal{I}} := & \min\limits_{x} & x^\top E x \\ & \text{s.t.} & \langle A^\top A, xx^\top\rangle - 2\bar{e}^\top x + p = 0, \\ & & \operatorname{diag}(xx^\top) - x = 0, \\ & & (A^\top A - I) \circ (xx^\top) = 0, \\ & & \bar{\mathcal{P}}_{\mathcal{I}}(xx^\top) \ge 0, \end{array} \tag{QQP$_{\text{SCP}}(\mathcal{I})$}$$

which is still an integer quadratic program (because of the constraint $\operatorname{diag}(xx^\top) - x = 0$).

9.2.2 SDP relaxation of the side chain positioning problem and its solvability

In this section, we derive the SDP relaxation of (QQP$_{\text{SCP}}(\mathcal{I})$), and show that the SDP relaxation is solvable.

Observe that (QQP$_{\text{SCP}}(\mathcal{I})$) is equivalent to:

$$\begin{array}{rll} v_{\text{scp}} = v_{\mathcal{I}} = & \min\limits_{x,X} & \langle E, X\rangle \\ & \text{s.t.} & \langle A^\top A, X\rangle - 2\bar{e}^\top x + p = 0, \\ & & \operatorname{diag}(X) - x = 0, \\ & & (A^\top A - I) \circ X = 0, \\ & & \bar{\mathcal{P}}_{\mathcal{I}}(X) \ge 0, \\ & & X = xx^\top. \end{array} \tag{9.12}$$

We obtain an SDP relaxation of (9.12) by replacing the constraint $X = xx^\top$ with $X \succeq xx^\top$. Since

$$\langle A^\top A, X\rangle - 2\bar{e}^\top x + p = \left\langle \begin{bmatrix} 1 & -\bar{e}^\top \\ -\bar{e} & A^\top A \end{bmatrix}, \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \right\rangle \quad \text{and} \quad X \succeq xx^\top \iff \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succeq 0,$$

defining

$$n := 1 + n_0, \quad \text{and} \quad A_{p,\lambda} := \begin{bmatrix} 1 & -\bar{e}^\top \\ -\bar{e} & A^\top A \end{bmatrix} \in \mathbb{S}^n, \tag{9.13}$$

we obtain

$$\begin{array}{rrl} v_{\text{scp}} = v_{\mathcal{I}} \geq d_{\mathcal{I}} := & \inf\limits_{x,X,Y} & \langle E, X\rangle \\ & \text{s.t.} & \left\langle A_{p,\lambda}, Y \right\rangle = 0, \\ & & \operatorname{diag}(X) - x = 0, \\ & & (A^\top A - I) \circ X = 0, \\ & & \bar{\mathcal{P}}_{\mathcal{I}}(X) \geq 0, \\ & & Y = \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succeq 0. \end{array} \tag{$\text{P}_{\text{SCP}}(\mathcal{I})$}$$

The SDP ($\text{P}_{\text{SCP}}(\mathcal{I})$) is in homogeneous equality form, i.e., other than the "normalization" constraint $Y_{11} = 1$, all the equality constraints on Y have zero on the right-hand side; in particular, results from [89, Section 3] applies.

Before stating the duality result regarding ($\text{P}_{\text{SCP}}(\mathcal{I})$), we remark that we can rewrite ($\text{P}_{\text{SCP}}(\mathcal{I})$) in a possibly more understandable form using linear maps

$$\text{arrow} \ : \ \mathbb{S}^n \to \mathbb{R}^{n-1} \ : \ \begin{bmatrix} \alpha & x^\top \\ x & X \end{bmatrix} \mapsto \operatorname{diag}(X) - x, \tag{9.14}$$

$$\text{bdiag}^{\tilde{m}} \ : \ \mathbb{S}^{\sum_k \tilde{m}_k} \to \mathbb{S}^{(\sum_k \tilde{m}_k)-1} \ : \ \begin{bmatrix} \alpha & x^\top \\ x & X \end{bmatrix} \mapsto ((A^{\tilde{m}})^\top A^{\tilde{m}} - I) \circ X, \tag{9.15}$$

which are defined for any dimension $n \geq 2$ and any integral vector $\tilde{m} \geq 0$. Same as for A, if $\tilde{m}$ is taken to be the default input m (from Page 130), then we simply write $\text{bdiag} := \text{bdiag}^m$. (These maps would be helpful for an easier understanding of the equivalent problem of ($\text{P}_{\text{SCP}}(\mathcal{I})$) that we will derive in Section 9.3.) The map arrow extracts the "arrow" part of a matrix, and the map bdiag zeros out the off-diagonal blocks of a matrix. The linear equation $\text{arrow}(Y) = 0$ ensures that the first column of Y equals the diagonal of Y, and the linear equation $\text{bdiag}(Y) = 0$ ensures

that the diagonal blocks of Y are diagonal matrices. In particular, ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) is equivalent to

$$\begin{aligned} d_{\mathcal{I}} = \quad & \inf_{x,X,Y} \quad \langle E, X\rangle \\ & \text{s.t.} \quad \left\langle A_{p,\lambda}, Y \right\rangle = 0, \\ & \qquad \mathrm{arrow}(Y) = 0, \\ & \qquad \mathrm{bdiag}(Y) = 0, \\ & \qquad \bar{\mathcal{P}}_{\mathcal{I}}(X) \geq 0, \\ & \qquad Y = \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succeq 0. \end{aligned}$$

We show that ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) is solvable using the duality theory. The Lagrangian of ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) is given by

$$\begin{aligned} \mathrm{L}(x, X, Y; \lambda, w, \Lambda, \eta, \Omega, \Phi) &= \langle E, X\rangle + \lambda\langle A_{p,\lambda}, Y\rangle + w^\top(\mathrm{diag}(X) - x) - \eta^\top(\bar{\mathcal{P}}_{\mathcal{I}}(X)) \\ &\quad + \langle \Lambda, (A^\top A - I) \circ X\rangle + \left\langle \Omega, Y - \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \right\rangle - \langle \Phi, Y\rangle \\ &= \langle E, X\rangle + \langle \lambda A_{p,\lambda} + \Omega - \Phi, Y\rangle + w^\top(\mathrm{diag}(X) - x) - \eta^\top(\bar{\mathcal{P}}_{\mathcal{I}}(X)) \\ &\quad + \langle \Lambda, (A^\top A - I) \circ X\rangle - \left\langle \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}, \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \right\rangle \\ &= -\Omega_{11} + \langle X, E + \mathrm{Diag}(w) + \Lambda \circ (A^\top A - I) - \bar{\mathcal{P}}^*_{\mathcal{I}}(\eta) - \Omega_{22}\rangle \\ &\quad - x^\top(w + 2\Omega_{21}) + \langle Y, \lambda A_{p,\lambda} - \Omega - \Phi\rangle, \end{aligned}$$

and the Lagrangian dual is given by

$$\begin{aligned} d^*_{\mathcal{I}} = \quad & \sup_{\substack{\lambda,w,\Lambda \\ \eta,\Omega,\Phi}} \big\{ -\Omega_{11} \ : \ \Omega_{22} = E + \mathrm{Diag}(w) + \Lambda \circ (A^\top A - I) - \bar{\mathcal{P}}^*_{\mathcal{I}}(\eta), \\ & \qquad\qquad 2\Omega_{21} + w = 0, \ \Phi = \lambda A_{p,\lambda} + \Omega, \ \Phi \succeq 0, \ \eta \geq 0 \big\}, \end{aligned}$$

which is equal to

$$\sup_{\substack{\lambda,w,\Lambda \\ \eta,\phi}} \left\{ -\phi \ : \ \lambda A_{p,\lambda} + \begin{bmatrix} \phi & -\frac{1}{2}w^\top \\ -\frac{1}{2}w & \big(E + \mathrm{Diag}(w) + \Lambda \circ (A^\top A - I) - \bar{\mathcal{P}}^*_{\mathcal{I}}(\eta)\big) \end{bmatrix} \succeq 0, \ \eta \geq 0 \right\}. \quad (9.16)$$

Observe that strong duality holds for the dual (9.16). Hence ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) has an optimal solution.

Proposition 9.2.1. *The Slater condition holds for* (9.16). *In particular,* ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) *and* (9.16) *have the same optimal value, and* ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) *has an optimal solution.*

Proof. Take

$$\hat{\lambda} = 0, \quad \hat{\eta} = \bar{e}, \quad \hat{w} = -\lambda_{\min}(E - \bar{\mathcal{P}}^*_{\mathcal{I}}(\hat{\eta}))\bar{e}, \quad \hat{\Lambda} = 0,$$

and any $\phi > 0$; then $\hat{\eta} > 0$ and

$$\begin{aligned}
&\hat{\lambda} A_{p,\lambda} + \begin{bmatrix} \hat{\phi} & -\frac{1}{2}\hat{w}^\top \\ -\frac{1}{2}\hat{w} & \left(E + \mathrm{Diag}(\hat{w}) + \hat{\Lambda} \circ (A^\top A - I) - \bar{\mathcal{P}}^*_{\mathcal{I}}(\hat{\eta})\right) \end{bmatrix} \\
= &\begin{bmatrix} \hat{\phi} & \frac{1}{2}\left(\lambda_{\min}(E - \bar{\mathcal{P}}^*_{\mathcal{I}}(\hat{\eta})) + 1\right)\bar{e}^\top \\ \frac{1}{2}\left(\lambda_{\min}(E - \bar{\mathcal{P}}^*_{\mathcal{I}}(\hat{\eta})) + 1\right)\bar{e} & E - \bar{\mathcal{P}}^*_{\mathcal{I}}(\hat{\eta}) - \lambda_{\min}(E - \bar{\mathcal{P}}^*_{\mathcal{I}}(\hat{\eta}))I + I \end{bmatrix}
\end{aligned}$$

is positive definite if $\hat{\phi}$ is large enough. Hence (9.16) satisfies the Slater condition.

On the other hand, ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) is feasible (since its rank-one feasible solutions correspond to the feasible solutions of ($\mathrm{IQP}_{\mathrm{SCP}}$)). Hence strong duality holds for (9.16), i.e., the optimal values of (9.16) and ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) are equal, and ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) has an optimal solution. □

9.3 Regularization of the SDP relaxation of (IQP)

We saw in Proposition 9.2.1 that ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) is solvable. However, the Slater condition indeed fails for ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$), meaning that we can regularize and reduce the problem size of ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$). In this section, we find the minimal face of ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) and regularize ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) by restricting its feasible region onto the minimal face.

We summarize the main result in Section 9.3.1. Section 9.3.2 contains the technical proofs for the result stated in Section 9.3.1, and can be skipped on the first reading.

9.3.1 Summary of the main result

The main result of this section is that the SDP ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) fails the Slater condition:

$$Y \text{ is feasible for } (\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})) \implies Y \in W\mathbb{S}^{n-p}_+W^\top \triangleleft \mathbb{S}^n_+,$$

for some full column rank matrix $W \in \mathbb{R}^{n\times(n-p)}$ (defined in (9.18)). In fact,

$$\mathrm{face}(\{Y \in \mathbb{S}^n : Y \text{ is feasible for } (\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))\ \}, \mathbb{S}^n_+) = W\mathbb{S}^{n-p}_+W^\top \triangleleft \mathbb{S}^n_+,$$

Moreover, there is no need to keep all the inequalities $Y_{ij} \geq 0$, as the diagonal blocks of Y are diagonal matrices (due to the equality constraint $(A^\top A - I) \circ Y_{2:n,\,2:n} = 0$). We can regularize ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) using these observations, arriving at an equivalent SDP that is smaller and more stable to solve.

Theorem 9.3.1. *The optimization problem* $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ *is equivalent to*

$$\begin{array}{rll} d_{\mathcal{I}} = & \inf\limits_{\hat{X}} & \langle \tilde{E}, \hat{X} \rangle \\ & \text{s.t.} & \hat{X}_{11} = 1, \\ & & \mathrm{arrow}(\hat{X}) = 0, \\ & & \mathrm{bdiag}^{m-\bar{e}}(\hat{X}) = 0, \\ & & \mathcal{P}_{\mathcal{I}}(W\hat{X}W^\top) \geq 0, \\ & & \hat{X} \succeq 0, \end{array} \qquad (\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$$

where $\tilde{E} := W^\top \begin{bmatrix} 0 & 0 \\ 0 & E \end{bmatrix} W \in \mathbb{S}^{n-p}$, $W \in \mathbb{R}^{n\times(n-p)}$ *is defined in* (9.18), $\mathrm{arrow} : \mathbb{S}^{n-p} \to \mathbb{R}^{n-p-1}$ *and* $\mathrm{bdiag}^{m-\bar{e}} : \mathbb{S}^{n-p} \to \mathbb{S}^{n-p-1}$ *are defined in* (9.14) *and* (9.15) *respectively. In fact,* Y *is feasible for* $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ *if and only if* $Y = W\hat{X}W^\top$ *for some feasible solution* $\hat{X}$ *of* $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$.

The dual of $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ *is given by*

$$\begin{array}{rll} \bar{d}^*_{\mathcal{I}} := & \sup\limits_{t,\lambda,\Lambda,\eta} & t \\ & \text{s.t.} & \bar{E} \succeq te_1e_1^\top + \mathrm{arrow}^*(\lambda) + (\mathrm{bdiag}^{m-\bar{e}})^*(\Lambda) + W^\top(\mathcal{P}^*_{\mathcal{I}}(\eta))W, \\ & & \eta \geq 0. \end{array} \qquad (\mathrm{D}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$$

If $\mathcal{I} \subseteq \mathcal{I}_{\geq 0}$, *where* $\mathcal{I}_{\geq 0}$ *is defined in* (9.10), *then both* $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ *and* $(\mathrm{D}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ *satisfy the Slater condition. In particular,* $d_{\mathcal{I}} = \bar{d}^*_{\mathcal{I}}$ *and both* $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ *and* $(\mathrm{D}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ *are solvable.*

Using Theorem 9.3.1, we can show that any rank-one solution of $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ gives an optimal solution to integer program $(\mathrm{IQP}_{\mathrm{SCP}})$ (see Corollary 9.3.5), and that $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ is equivalent to the SDP relaxation (9.5) introduced in [25] when $\mathcal{I} = \mathcal{I}_{\geq 0}$ (see Proposition 9.3.7), though thanks to the regularization, $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ has a smaller matrix variable and less constraints.

9.3.2 Proof of the main results

In this section, we provide the details for the proof of Theorem 9.3.1.

It is quite easy to see that $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ fails the Slater condition, simply by noticing that the constraint $\langle A_{p,\lambda}, Y\rangle = 0$ involves a positive semidefinite matrix $A_{p,\lambda}$ (so $\langle A_{p,\lambda}, Y\rangle = 0$ and $Y \succeq 0$ imply that $Y \in \mathbb{S}^n_+ \cap \{A_{p,\lambda}\}^\perp \triangleleft \mathbb{S}^n_+$). Before we prove that $A_{p,\lambda}$ is positive semidefinite, we define the matrices

$$B_k := \begin{bmatrix} I_{k-1} \\ -\bar{e}_{k-1}^\top \end{bmatrix} \in \mathbb{R}^{k\times(k-1)} \qquad (9.17)$$

for any positive integer k, and

$$W := \begin{array}{c} \\ 1 \\ m_1 \\ m_2 \\ \\ m_p \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & m_1-1 & m_2-1 & & m_p-1 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ e_{m_1} & B_{m_1} & 0 & \cdots & 0 \\ e_{m_2} & 0 & B_{m_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ e_{m_p} & 0 & 0 & \cdots & B_{m_p} \end{bmatrix} \end{array} \in \mathbb{R}^{n\times(n-p)} \quad \text{and} \quad w := W_{:1} = \begin{bmatrix} 1 \\ e_{m_1} \\ e_{m_2} \\ \vdots \\ e_{m_p} \end{bmatrix} \in \mathbb{R}^n. \tag{9.18}$$

(In particular, B_1 is a vacuous matrix.) B_k is the nullspace representation of $\bar{e}_k^\top$:

$$\ker(\bar{e}_k^\top) = \{\bar{e}_k\}^\perp = \operatorname{range}(B_k).$$

Proposition 9.3.2. *[24, Lemma 1] The matrix $A_{p,\lambda} \in \mathbb{S}^n$ defined in* (9.13) *is positive semidefinite, and*

$$\ker(A_{p,\lambda}) = \operatorname{span}\left(\{w\} \cup \left\{ \begin{bmatrix} 0 \\ q^{(1)} \\ \vdots \\ q^{(p)} \end{bmatrix} : \bar{e}^\top q^{(k)} = 0,\ q^{(k)} \in \mathbb{R}^{m_k},\ \forall\, k \in 1:p \right\} \right) = \operatorname{range}(W).$$

In particular, $\dim(\ker(A_{p,\lambda})) = n - p$.

Proof. The Schur complement of $A_{p,\lambda}$ (with respect to its (1,1)-entry $(A_{p,\lambda})_{11} = p$) is given by

$$A^\top A - \frac{1}{p}\bar{e}\bar{e}^\top \in \mathbb{S}^{n_0},$$

which is positive semidefinite if and only if $A_{p,\lambda}$ is positive semidefinite. In fact, for any $x = \left(x^{(1)}; x^{(2)}; \ldots; x^{(p)}\right) \in \mathbb{R}^{m_1+m_2+\ldots+m_p}$, we have

$$(\bar{e}^\top x)^2 = \left(\sum_{k=1}^p \bar{e}^\top x^{(k)}\right)^2 \le \left(\sum_{k=1}^p |\bar{e}^\top x^{(k)}|\right)^2 \le p\sum_{k=1}^p |\bar{e}^\top x^{(k)}|^2 = p\|Ax\|^2,$$

i.e., $x^\top(A^\top A)x \ge \frac{1}{p}x^\top\bar{e}^\top\bar{e}x$. Therefore $A^\top A - \frac{1}{p}\bar{e}\bar{e}^\top$ is positive semidefinite, implying that $A_{p,\lambda}$ is positive semidefinite too.

Now we prove the second equality in (9.18). For each $k \in 1:p$ and any $q^{(k)} \in \mathbb{R}^{m_k}$, $\bar{e}^\top q^{(k)} = 0$ if and only if $q_k^{(\in)} \operatorname{range}(B_{m_k})$. For any $q = (q^{(1)}; q^{(2)}; \cdots; q^{(p)}) \in \mathbb{R}^{m_1+m_2+\cdots+m_p}$, $\bar{e}^\top q^{(k)} = 0$ for all $k \in 1:p$ if and only if $q^{(k)} \in \operatorname{range}(B_{m_k})$ for all $k \in 1:p$, if and only q lies in the range of the block diagonal matrix $\operatorname{Diag}(B_{m_1}, B_{m_2}, \ldots, B_{m_p})$. This proves the second equality.

For the first equality, it is immediate that the columns of W lie in $\ker(A_{p,\lambda})$, so it suffices to show that $\dim(\ker(A_{p,\lambda})) \leq \operatorname{rank}(W)$. First note that W is of full column rank, i.e., $\operatorname{rank}(W) = n - p$. Define $u^{(k)} := (-1; 0; \ldots; \bar{e}_{m_k}^\top; \ldots; 0) \in \mathbb{R}^n$ for $k \in 1 : p$. Then $\left\{u^{(1)}, \ldots, u^{(p)}\right\} \in \operatorname{range}(A_{p,\lambda})$ is linearly independent. This implies that $\operatorname{rank}(A_{p,\lambda}) \geq p$, and $\dim(\ker(A_{p,\lambda})) \leq n - p = \operatorname{rank}(W)$. (In particular, the columns of W form a basis of $\ker(A_{p,\lambda})$.) $\square$

A consequence of Proposition 9.3.2 is that the feasible region of $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ is contained in $W\mathbb{S}_+^{n-p}W^\top$. In fact, we can show that $W\mathbb{S}_+^{n-p}W^\top$ is the minimal face of $\mathbb{S}_+^n$ containing the feasible region of $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$. (See Theorem 9.3.6 for the formal proof.) To prove this claim, we use the smaller face $W\mathbb{S}_+^{n-p}W^\top$ to derive an SDP equivalent to $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ (see Proposition 9.3.4 for the proof of equivalence):

$$\begin{aligned} d_{\mathcal{I}} = \quad & \inf_{\hat{X}} \quad \left\langle \begin{bmatrix} 0 & 0 \\ 0 & E \end{bmatrix}, W\hat{X}W^\top \right\rangle \\ & \text{s.t.} \quad (W\hat{X}W^\top)_{11} = 1, \\ & \qquad \langle A_j, W\hat{X}W^\top \rangle = 0, \ \forall\, j \in 1 : n_0, \\ & \qquad \langle e_{i+1,\,j+1} + e_{j+1,\,i+1}, W\hat{X}W^\top \rangle = 0, \ \forall\, (i,j) \in \mathcal{B}, \\ & \qquad \mathcal{P}_{\mathcal{I}}(W\hat{X}W^\top) \geq 0, \\ & \qquad \hat{X} \succeq 0, \end{aligned} \tag{9.19}$$

where

- for $j \in 1 : n_0$, $A_j \in \mathbb{S}^n$ is defined as

$$A_j \ := \ e_{j+1}e_{j+1}^\top - \frac{1}{2}\left(e_1 e_{j+1}^\top + e_{j+1}e_1^\top\right) \ = \ \begin{bmatrix} 0 & -\frac{1}{2}e_j^\top \\ -\frac{1}{2}e_j & e_j e_j^\top \end{bmatrix}, \tag{9.20}$$

  so that

$$\operatorname{diag}(X) = x \iff \left\langle A_j, \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \right\rangle = 0, \ \forall\, j \in 1 : n_0; \tag{9.21}$$

- $\mathcal{B}$ is defined in (9.8), so that for all $Y = \begin{bmatrix} \alpha & x^\top \\ x & X \end{bmatrix} \in \mathbb{S}^n$,

$$(A^\top A - I) \circ X = 0 \iff X_{ij} = 0, \ \forall\, (i,j) \in \mathcal{B} \iff Y_{i+1,\,j+1} = 0, \ \forall\, (i,j) \in \mathcal{B}; \tag{9.22}$$

- $\mathcal{P}_{\mathcal{I}}$ is the projection defined as

$$\mathcal{P}_{\mathcal{I}} : \mathbb{S}^n \to \mathbb{R}^{|\mathcal{I}|} : \begin{bmatrix} \alpha & x^\top \\ x & X \end{bmatrix} \mapsto \bar{\mathcal{P}}_{\mathcal{I}}(X),$$

(where $\bar{\mathcal{P}}_{\mathcal{I}}$ is defined in (9.11)), so that for all $Y = \begin{bmatrix} \alpha & x^\top \\ x & X \end{bmatrix} \in \mathbb{S}^n$,

$$\mathcal{P}_{\mathcal{I}}(Y) = 0 \iff \bar{\mathcal{P}}_{\mathcal{I}}(X) = 0. \tag{9.23}$$

We can simplify (9.19) by computing $W^\top A_{p,\lambda} W$, $W^\top A_j W$ (for $j \in 1 : n_0$) and $W^\top (e_{i+1} e_{j+1}^\top) W$ (for $(i,j) \in \mathcal{B}$).

Lemma 9.3.3. *[24, Lemma 2] For $j \in 1 : n_0$ and A_j defined in* (9.20) *(and recalling the definition of $\bar{m}_k$ and $\mathcal{V}_k$ for $k \in 1 : p$ on Page 130):*

(1) if $j \in \mathcal{V}_k$ and $j \neq \bar{m}_k$ (for some unique $k \in 1 : p$), then

$$W^\top A_j W = e_{j-k+2}\,(e_{j-k+2})^\top - \frac{1}{2}(e_1 (e_{j-k+2})^\top + e_{j-k+2}\, e_1^\top) = \begin{bmatrix} 0 & -\frac{1}{2}(e_{j-k+1})^\top \\ \frac{1}{2} e_{j-k+1} & e_{j-k+1}(e_{j-k+1})^\top \end{bmatrix} \in \mathbb{S}^{n-p};$$

(2) if $j = \bar{m}_k$ for some $k \in 1 : p$, then

$$W^\top A_j W = W^\top A_{\bar{m}_k} W = \begin{bmatrix} 0 & 0 & \cdots & -\frac{1}{2}\bar{e}^\top & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ -\frac{1}{2}\bar{e} & 0 & \cdots & \bar{E} & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}. \tag{9.24}$$

For all $(i,j) \in \mathcal{B}$, i.e., $(i,j) \in \mathcal{V}_k$ with $i < j$:

(3) if $j < \bar{m}_k$, then

$$W^\top (e_{i+1}(e_{j+1})^\top + e_{j+1}(e_{i+1})^\top) W = e_{i-k+2}(e_{j-k+2})^\top + e_{j-k+2}(e_{i-k+2})^\top; \tag{9.25}$$

(4) if $j = \bar{m}_k$, then $W^\top (e_{i+1}(e_{j+1})^\top + e_{j+1}(e_{i+1})^\top) W$ equals

$$\begin{aligned} & W^\top (e_{i+1}(e_{\bar{m}_k+1})^\top + e_{\bar{m}_k+1}(e_{i+1})^\top) W \\ = & \begin{bmatrix} 0 & 0 & \cdots & (e_{i-\bar{m}_{k-1}})^\top & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ e_{i-\bar{m}_{k-1}} & 0 & \cdots & -e_{i-\bar{m}_{k-1}}\bar{e}^\top - \bar{e}\,(e_{i-\bar{m}_{k-1}})^\top & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}. \end{aligned} \tag{9.26}$$

Proof. We first write down $W^\top e_j$ for $j \in 1:n$. By definition of W in (9.18), $W^\top e_1 \in \mathbb{R}^{n-p}$. For each $j \in 1:n_0$, there exists a unique $k \in 1:p$ such that $j \in \mathcal{V}_k = (\bar{m}_{k-1}+1):\bar{m}_k$. Then $j - \bar{m}_{k-1} \in 1:m_k$, and

$$e_{j+1}^\top W = W_{j,:} = \begin{cases} \begin{bmatrix} \overset{1}{0} & \overset{m_1-1}{0} & \cdots & \overset{m_k-1}{e_{j-\bar{m}_{k-1}}^\top} & \cdots & \overset{m_p-1}{0} \end{bmatrix} & \text{if } j \neq \bar{m}_k, \\ \begin{bmatrix} \overset{1}{1} & \overset{m_1-1}{0} & \cdots & \overset{m_k-1}{-\bar{e}^\top} & \cdots & \overset{m_p-1}{0} \end{bmatrix} & \text{if } j = \bar{m}_k, \end{cases} \tag{9.27}$$

i.e.,

$$W^\top e_{j+1} = \begin{cases} e_{j-k+2} & \text{if } j \neq \bar{m}_k, \\ e_1 + \sum_{i \in \mathcal{V}_k \setminus \{\bar{m}_k\}} e_{i-k+1} & \text{if } j = \bar{m}_k. \end{cases} \tag{9.28}$$

Now for each $j \in 1:n_0$, we compute $W^\top A_j W$. Let $k \in 1:p$ be such that $j \in \mathcal{V}_k$.

- If $j \neq \bar{m}_k$, then

$$\begin{aligned} W^\top A_j W &= W^\top (e_{j+1}e_{j+1}^\top) W - \frac{1}{2} W^\top (e_1 e_{j+1}^\top + e_{j+1} e_1^\top) W \\ &= e_{j-k+2} e_{j-k+2}^\top - \frac{1}{2}\left(e_1 e_{j-k+2}^\top + e_{j-k+2} e_1^\top \right) \\ &= \begin{bmatrix} 0 & -\frac{1}{2} e_{j-k+1}^\top \\ \frac{1}{2} e_{j-k+1} & e_{j-k+1} e_{j-k+1}^\top \end{bmatrix} \in \mathbb{S}^{n-p}. \end{aligned}$$

- If $j = \bar{m}_k$, then by (9.27),

$$\begin{aligned} W^\top A_j W &= W^\top A_{\bar{m}_k} W \\ &= \begin{bmatrix} 1 & 0 & \cdots & -\bar{e}^\top & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ -\bar{e} & 0 & \cdots & \bar{E} & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix} - \begin{bmatrix} 1 & 0 & \cdots & -\frac{1}{2}\bar{e}^\top & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ -\frac{1}{2}\bar{e} & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}, \end{aligned}$$

  so (9.24) holds.

Now we compute $W^\top (e_{i+1}e_{j+1}^\top + e_{j+1}e_{i+1}^\top) W$. If $i < j < \bar{m}_k$, then (9.25) follows immediately from (9.28). If $i < j = \bar{m}_k$, then (9.26) follow from (9.27). □

The computations in Lemma 9.3.3 allow us to simplify the facially reduced program (9.19). We restate and prove the first part of Theorem 9.3.1 below.

Proposition 9.3.4. *[24, Theorem 1] The optimization problem* $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ *is equivalent to the regularized problem* $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$*, which we restate here:*

$$\begin{aligned} d_{\mathcal{I}} = \quad & \inf_{\hat{X}} \quad \langle \tilde{E}, \hat{X} \rangle \\ & \text{s.t.} \quad \hat{X}_{11} = 1, \\ & \qquad \operatorname{arrow}(\hat{X}) = 0, \\ & \qquad \operatorname{bdiag}^{m-\bar{e}}(\hat{X}) = 0, \\ & \qquad \mathcal{P}_{\mathcal{I}}(W\hat{X}W^\top) \geq 0, \\ & \qquad \hat{X} \succeq 0, \end{aligned} \tag{9.29}$$

where $\tilde{E} := W^\top \begin{bmatrix} 0 & 0 \\ 0 & E \end{bmatrix} W \in \mathbb{S}^{n-p}$*,* $\operatorname{arrow} : \mathbb{S}^{n-p} \to \mathbb{R}^{n-p-1}$ *and* $\operatorname{bdiag}^{m-\bar{e}} : \mathbb{S}^{n-p} \to \mathbb{S}^{n-p-1}$ *are defined in* (9.14) *and* (9.15) *respectively. In fact,* Y *is feasible for* $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ *if and only if* $Y = W\hat{X}W^\top$ *for some feasible solution* $\hat{X}$ *of* $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$*.*

Proof. First we prove the earlier claim that $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ and (9.19) are equivalent.

Note that since $A_{p,\lambda} \succeq 0$ with $\ker(A_{p,\lambda}) = \operatorname{range}(W)$ (see Proposition 9.3.2),

$$\langle A_{p,\lambda}, Y \rangle = 0, Y \succeq 0 \iff A_{p,\lambda} Y = 0, Y \succeq 0 \iff Y \in W\mathbb{S}^{n-p}_+ W^\top.$$

Hence we can replace Y with $W\hat{X}W^\top$ in $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$. Since

$$(W\hat{X}W^\top)_{11} = e_1^\top W\hat{X}W^\top e_1 = e_1^\top \hat{X} e_1,$$

by (9.21), (9.22) and (9.23), the SDPs $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ and (9.19) are equivalent.

By Lemma 9.3.3, for $k \in 1:p$,

$$W^\top A_{\bar{m}_k} W - \sum_{j \in \mathcal{V}_k \setminus \{\bar{m}_k\}} W^\top A_j W = \begin{bmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & \bar{E} - I & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{bmatrix}$$

and $W^\top(e_{i+1}(e_{\bar{m}_k+1})^\top + e_{\bar{m}_k+1}(e_{i+1})^\top)W$ are linear combinations of

$$\begin{bmatrix} 0 & -\frac{1}{2}(e_{j-k+1})^\top \\ \frac{1}{2}e_{j-k+1} & e_{j-k+1}(e_{j-k+1})^\top \end{bmatrix} \quad \text{and} \quad e_{i-k+2}(e_{j-k+2})^\top + e_{j-k+2}(e_{i-k+2})^\top$$

(for $i,j \in \mathcal{V}_k$ with $i < j < \bar{m}_k$). Hence $\langle A_j, W\hat{X}W^\top\rangle = 0$ for all $j \in 1:n_0$ and $\langle e_{i+1,\,j+1} + e_{j+1,\,i+1}, W\hat{X}W^\top\rangle = 0$ for all $(i,j) \in \mathcal{B}$ if and only for all $k \in 1:p$ and $i,j \in \mathcal{V}_k$ with $i < j < \bar{m}_k$,

$$\langle A_j, W\hat{X}W^\top\rangle = 0 \quad \text{and} \quad \langle e_{i+1}\, e_{j+1}^\top + e_{j+1}\, e_{i+1}^\top, W\hat{X}W^\top\rangle = 0,$$

if and only if for all $k \in 1:p$ and $i,j \in \mathcal{V}_k$ with $i < j < \bar{m}_k$,

$$\left\langle \begin{bmatrix} 0 & -\frac{1}{2}e_i^\top \\ -\frac{1}{2}e_i & e_i e_i^\top \end{bmatrix}, \hat{X} \right\rangle = 0 \quad \text{and} \quad \langle e_{i-k+2}(e_{j-k+2})^\top + e_{j-k+2}(e_{i-k+2})^\top, \hat{X}\rangle = 0,$$

if and only if

$$\operatorname{arrow}(\hat{X}) = 0 \quad \text{and} \quad \operatorname{bdiag}^{(m_1-1,m_2-1,\ldots,m_p-1)}(\hat{X}) = 0.$$

Therefore, Y is feasible for $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ if and only if $Y = W\hat{X}W^\top$ for some feasible solution $\hat{X}$ of $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$. Moreover, Y and $\hat{X}$ have the same objective value in $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ and $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ respectively. Consequently, (9.19) and (9.29) (i.e., $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$) are equivalent. □

We now prove the immediate result that any rank-one optimal solution of (9.29) gives an optimal solution of the integer program $(\mathrm{IQP}_{\mathrm{SCP}})$.

Corollary 9.3.5. *If $\hat{X} = xx^\top$ is an optimal solution of (9.29) with $x_1 = 1$, then $y := Wx$ satisfies $y_1 = 1$ and $y_{2:n}$ is an optimal solution of $(\mathrm{IQP}_{\mathrm{SCP}})$.*

Proof. By Proposition 9.3.4, $yy^\top$ is an optimal solution of $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$. Then $y_{2:n}$ is a feasible solution of $(\mathrm{QQP}_{\mathrm{SCP}}(\mathcal{I}))$ (and of $(\mathrm{IQP}_{\mathrm{SCP}})$), and

$$v_{\mathrm{scp}} \le \langle E, y_{2:n}y_{2:n}^\top\rangle = d_{\mathcal{I}} \le v_{\mathrm{scp}}$$

implies that $y_{2:n}$ is an optimal solution of $(\mathrm{QQP}_{\mathrm{SCP}}(\mathcal{I}))$ and of $(\mathrm{IQP}_{\mathrm{SCP}})$. □

Now we write down the dual of (9.29), and show that both (9.29) and its dual satisfy the Slater condition. The Lagrangian of (9.29) is given by

$$\begin{aligned}\mathrm{L}(\hat{X}; t, \lambda, \Lambda, \eta, \Omega) &= \langle \tilde{E}, \hat{X}\rangle - \lambda^\top(\operatorname{arrow}(\hat{X})) + t(1 - e_1^\top \hat{X} e_1) \\ &\qquad - \langle \Lambda, \operatorname{bdiag}^{m-\bar{e}}(\hat{X})\rangle - \eta^\top(\mathcal{P}_{\mathcal{I}}(W\hat{X}W^\top)) - \langle \Omega, \hat{X}\rangle \\ &= t + \langle X, \tilde{E} - te_1e_1^\top - \operatorname{arrow}^*(\lambda) - (\operatorname{bdiag}^{m-\bar{e}})^*(\Lambda) - W^\top(\mathcal{P}^*_{\mathcal{I}}(\eta))W - \Omega\rangle;\end{aligned}$$

hence the dual of (9.29) is given by $(\mathrm{D}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$, which we restate here:

$$\begin{aligned} \bar{d}^*_{\mathcal{I}} := \quad \sup_{t,\lambda,\Lambda,\eta} \quad & t \\ \text{s.t.} \quad & \bar{E} \succeq te_1e_1^\top + \operatorname{arrow}^*(\lambda) + (\operatorname{bdiag}^{m-\bar{e}})^*(\Lambda) + W^\top(\mathcal{P}^*_{\mathcal{I}}(\eta))W, \\ & \eta \ge 0. \end{aligned} \tag{9.30}$$

This proves the second claim of Theorem 9.3.1.

Now we prove the last claim of Theorem 9.3.1: we show that both (9.29) and its dual (9.30) satisfies the Slater condition.

Theorem 9.3.6. *[24, Proposition 5] Let $\mathcal{I} \subseteq \mathcal{I}_{\geq 0}$, where $\mathcal{I}_{\geq 0}$ is defined in* (9.10) *Then both* (9.29) *and* (9.30) *satisfy the Slater condition. In particular, $d_{\mathcal{I}} = \bar{d}^*_{\mathcal{I}}$, and both* (9.29) *and* (9.30) *are solvable.*

Proof. We first prove that (9.29) satisfies the Slater condition. Define

$$\hat{\hat{X}} := e_1 e_1^\top + \frac{1}{2(n_0-p)} \operatorname{arrow}^*(\bar{e}_{n_0-p}) \in \mathbb{S}^{n-p},$$
$$\hat{D} := \bar{E} - e_1 e_1^\top - \operatorname{arrow}^*(\bar{e}) - (\operatorname{bdiag}^{m-\bar{e}})^*(\bar{E}) \in \mathbb{S}^{n-p}.$$

Then

$$\hat{\hat{X}}_{11} = 1, \quad \operatorname{arrow}(\hat{\hat{X}}) = 0, \quad \operatorname{bdiag}^{m-\bar{e}}(\hat{\hat{X}}) = 0,$$
$$\hat{D}_{11} = 0, \quad \operatorname{arrow}(\hat{D}) = 0, \quad \operatorname{bdiag}^{m-\bar{e}}(\hat{\hat{X}}) = 0.$$

In particular, $\hat{\hat{X}} + \alpha \hat{D}$ satisfies the linear equality constraints in (9.29) for all $\alpha \in \mathbb{R}$.

Next we show that $\hat{\hat{X}} + \alpha \hat{D}$ is positive definite for sufficiently small $\alpha > 0$. It is immediate that $\hat{\hat{X}} \succ 0$: the Schur complement of $\hat{\hat{X}}$ with respect to $\hat{\hat{X}}_{11} = 1$ is

$$\begin{aligned}\frac{1}{2(n_0-p)} I_{n_0-p} - \frac{1}{4(n_0-p)^2} \bar{E}_{n_0-p} &\succeq \frac{1}{2(n_0-p)} I_{n_0-p} - \frac{1}{4(n_0-p)^2} \lambda_{\min}(\bar{E}_{n_0-p}) I_{n_0-p} \\ &= \frac{1}{4(n_0-p)} I_{n_0-p} \succ 0,\end{aligned}$$

since $\lambda_{\min}(\bar{E}_{n_0-p}) = n_0 - p$. On the other hand, since $\hat{D} \neq 0$ has a zero diagonal, $\hat{D}$ is indefinite. For any $0 < \alpha < \frac{\lambda_{\min}(\hat{\hat{X}})}{-\lambda_{\min}(\hat{D})}$,

$$\lambda_{\min}(\hat{\hat{X}} + \alpha \hat{D}) \geq \lambda_{\min}(\hat{\hat{X}}) + \alpha \lambda_{\min}(\hat{D}) > 0,$$

i.e., $\hat{\hat{X}} + \alpha \hat{D}$ is positive definite.

Now we show that for sufficiently small $\alpha > 0$, $\mathcal{P}_{\mathcal{I}}(W(\hat{\hat{X}} + \alpha \hat{D})W^\top) > 0$. Write

$$\hat{Y} := W \hat{\hat{X}} W^\top = \begin{array}{c} \\ 1 \\ m_1 \\ m_2 \\ \\ m_p \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & m_1 & m_2 & & m_p \end{array} \\ \begin{bmatrix} 1 & \hat{Y}^{01} & \hat{Y}^{02} & \dots & \hat{Y}^{0p} \\ \hat{Y}^{10} & \hat{Y}^{11} & \hat{Y}^{12} & \dots & \hat{Y}^{1p} \\ \hat{Y}^{20} & \hat{Y}^{21} & \hat{Y}^{22} & \dots & \hat{Y}^{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \hat{Y}^{p0} & \hat{Y}^{p1} & \hat{Y}^{p2} & \dots & \hat{Y}^{pp} \end{bmatrix} \end{array};$$

then for $k \in 1:p$,

$$\hat{Y}^{k0} = \begin{bmatrix} \frac{1}{2(n_0-p)}\bar{e} \\ 1 - \frac{m_k-1}{2(n_0-p)} \end{bmatrix}, \quad \hat{Y}^{kk} = \operatorname{Diag}(\hat{Y}_{k0}),$$

$$\hat{Y}^{kl} = \begin{array}{c} \\ m_k \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} m_l-1 & 1 \end{array} \\ \begin{bmatrix} 0 & \frac{1}{2(n_0-p)}\bar{e} \\ \frac{1}{2(n_0-p)}\bar{e}^\top & 1 - \frac{m_k+m_l-2}{2(n_0-p)} \end{bmatrix} \end{array}, \ \forall, l \neq k.$$

Next, write

$$\tilde{Y} := W\hat{D}W^\top = \begin{array}{c} \\ 1 \\ m_1 \\ m_2 \\ \\ m_p \end{array} \begin{array}{c} \begin{array}{ccccc} 1 & m_1 & m_2 & & m_p \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \tilde{Y}^{12} & \cdots & \tilde{Y}^{1p} \\ 0 & \tilde{Y}^{21} & 0 & \cdots & \tilde{Y}^{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \tilde{Y}^{p1} & \tilde{Y}^{p2} & \cdots & 0 \end{bmatrix} \end{array},$$

where

$$\tilde{Y}^{kl} = \begin{array}{c} \\ m_k-1 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} m_l-1 & 1 \end{array} \\ \begin{bmatrix} \bar{E} & -(m_l-1)\bar{e} \\ -(m_k-1)\bar{e}^\top & (m_k-1)(m_l-1) \end{bmatrix} \end{array}, \ \forall\, k \neq l.$$

Then for any $0 < \alpha < \frac{1}{2(n_0-p)^2}$, for all $k \neq l$,

$$\hat{Y}^{kl} + \alpha\tilde{Y}^{kl} = \begin{array}{c} \\ m_k-1 \\ 1 \end{array} \begin{array}{c} \begin{array}{cc} m_l-1 & 1 \end{array} \\ \begin{bmatrix} \alpha\bar{E} & \left(\frac{1}{2(n_0-p)} - \alpha(m_l-1)\right)\bar{e} \\ \left(\frac{1}{2(n_0-p)} - \alpha(m_k-1)\right)\bar{e}^\top & 1 - \frac{m_k+m_l-2}{2(n_0-p)} + \alpha(m_k-1)(m_l-1) \end{bmatrix} \end{array} > 0,$$

since $m_k - 1 \leq n_0 - p$. Therefore $(W(\hat{\hat{X}} + \alpha\hat{D})W^\top)_{ij} = (\hat{Y} + \alpha\tilde{Y})_{ij} > 0$ for all $(i-1, j-1) \in \mathcal{I}_{\geq 0}$. In particular, $\mathcal{P}_{\mathcal{I}}(\hat{\hat{X}} + \alpha\hat{D}) > 0$.

Consequently, for sufficiently small $\alpha > 0$, $\hat{\hat{X}} + \alpha\hat{D}$ is a Slater point for (9.29).

Now for (9.30), take

$$\hat{\eta} = \bar{e}, \quad \hat{\Omega} = 0, \quad \hat{\lambda} = \left(\lambda_{\min}(\tilde{E} - W^\top(\mathcal{P}^*_{\mathcal{I}}(\eta))W) - 1\right)\bar{e}, \quad \hat{t} > \frac{1}{4}\|\hat{\lambda}\|^2;$$

then

$$\begin{aligned}
&\tilde{E} - \hat{t} - \operatorname{arrow}^*(\hat{\lambda}) - (\operatorname{bdiag}^{m-\bar{e}})^*(\Lambda) - W^\top(\mathcal{P}_{\mathcal{I}}^*(\eta))W \\
&= (\tilde{E} - W^\top(\mathcal{P}_{\mathcal{I}}^*(\eta))W - \lambda_{min}(\tilde{E} - W^\top(\mathcal{P}_{\mathcal{I}}^*(\eta))W)I) + \begin{bmatrix} \hat{t} & \frac{1}{2}\hat{\lambda} \\ \frac{1}{2}\hat{\lambda} & I \end{bmatrix} \\
&\succeq \begin{bmatrix} \hat{t} & \frac{1}{2}\hat{\lambda} \\ \frac{1}{2}\hat{\lambda} & I \end{bmatrix},
\end{aligned}$$

which is positive definite because its Schur complement with respect to the (2,2)-block, $t - \frac{1}{4}\|\hat{\lambda}\|^2$, is positive. In addition, $\hat{\eta} > 0$, so $(\hat{t}, \hat{\lambda}, \hat{\Lambda}, \hat{\eta})$ is a Slater point for (9.30). Consequently, both (9.29) and (9.30) satisfy the Slater condition. □

9.3.3 Equivalence to the SDP relaxation by Chazelle *et al.*

In this section, we show that the SDP relaxations $(\mathrm{P}_{\mathrm{SCP}}^{\mathrm{reg}}(\mathcal{I}))$ and the SDP relaxation (9.5) proposed by Chazelle *et al.* [25] are equivalent when $\mathcal{I} = \mathcal{I}_{\geq 0}$.

Proposition 9.3.7. *[24, Corollary 1] If $\mathcal{I} = \mathcal{I}_{\geq 0}$, then $(\mathrm{P}_{\mathrm{SCP}}^{\mathrm{reg}}(\mathcal{I}))$ is equivalent to (9.5).*

Proof. We show that the feasible regions of $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ and of (9.5) are the same.

Let Y be a feasible solution of $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ (and of (9.19)). Then $Y = W\hat{X}W^\top$ for some feasible solution $\hat{X}$ of $(\mathrm{P}_{\mathrm{SCP}}^{\mathrm{reg}}(\mathcal{I}))$ by Proposition 9.3.4. Also, $Y_{11} = 0$ and $\operatorname{arrow}(Y) = 0$, so the third and fourth constraints of (9.5) are satisfied. To see that the second constraint of (9.5) is satisfied by Y, simply note that for each $k \in 1:p$,

$$\sum_{i,j \in \mathcal{V}_k} Y_{i+1,j+1} = \left\langle W^\top \left(\sum_{i \in \mathcal{V}_k} e_i \right) \left(\sum_{i \in \mathcal{V}_k} e_i \right)^\top W, \hat{X} \right\rangle = \langle e_1 e_1^\top, \hat{X} \rangle = 1.$$

This together with $(A^\top A - I) \circ Y_{2:n,\,2:n} = 0$ (implying that the diagonal blocks of $Y_{2:n,\,2:n}$ are diagonal) and the arrow constraint means that Y also satisfies the first constraint.

Finally, since the diagonal blocks are nonnegative and $\mathcal{P}_{\mathcal{I}_{\geq 0}}(Y) \geq 0$ (implying that $Y_{i+1,j+1} \geq 0$ whenever $(i,j) \notin \mathcal{B}$), we have that $Y \geq 0$. Therefore, if Y is feasible for $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$, then Y is also feasible for (9.5).

Conversely, suppose that Y is feasible for (9.5). We show that Y is feasible for $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$. It suffices to check that $\langle A_{p,\lambda}, Y \rangle = 0$ and $(A^\top A - I) \circ Y_{2:n,\,2:n} = 0$.

- Note that $Y \geq 0$ together and

$$\sum_{j \in \mathcal{V}_k} Y_{j+1,\, j+1} = 1 \sum_{i,j \in \mathcal{V}_k} Y_{i+1,\, j+1}, \quad \forall\, k \in 1:p$$

  imply that $Y_{i+1,\, j+1} = 0$ for all distinct $i, j \in \mathcal{V}_k$, for each $k \in 1:p$. Therefore $(A^\top A - I) \circ Y_{2:n,\, 2:n} = 0$.

- Since $(A^\top A) \circ Y_{2:n,\, 2:n} = I \circ Y$, we have $\langle A^\top A, Y \rangle = \operatorname{tr}(Y) - 1$. On the other hand, the arrow constraint implies that $\sum_{k=1}^{p} \sum_{j \in \mathcal{V}_k} Y_{j+1,1} = \operatorname{tr}(Y) - 1 = \bar{e}^\top Y_{2:n,1} = p$. Therefore

$$\langle A_{p,\lambda}, Y \rangle = \langle A^\top A, Y_{2:n,\, 2:n} \rangle - 2\bar{e}^\top Y_{2:n,\,1} + p = 0.$$

Therefore Y is feasible for $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$. Hence $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ and (9.5) have the same feasible region. They also have the same objective, so they are equivalent. Finally, since $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ and $(\mathrm{P}_{\mathrm{SCP}}(\mathcal{I}))$ are equivalent, $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ and (9.5) are equivalent as well. □

In particular, Proposition 9.3.7 indicates that (9.5) fails the Slater condition as well: the matrix variable Y can be restricted onto the proper face $W\mathbb{S}^{n-p}_{+}W^\top$ of $\mathbb{S}^n_+$ and some of the inequalities $Y_{ij} \geq 0$ can indeed be replaced by equality.

9.4 Implementation: obtaining an optimal solution of (IQP)

In this section, we discuss some implementation issues of obtaining near optimal solutions of the integer program $(\mathrm{IQP}_{\mathrm{SCP}})$ from the SDP relaxation $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$.

- In Section 9.4.1, we describe the cutting plane technique: we solve $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ multiple times; after each iteration we add a "sensible" amount of indices for "useful" nonnegativity constraints to $\mathcal{I}$, in order to obtain a tighter SDP relaxation for the next iteration and at the same time keep the computational costs in check.

- In Section 9.4.2 we review some existing techniques for "rounding" a feasible solution of $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$, i.e., obtaining from a feasible solution of $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ a feasible integral solution of the integer program $(\mathrm{IQP}_{\mathrm{SCP}})$.

- In Section 9.5.1, we discuss some measures of the quality of feasible integral solutions of $(\mathrm{IQP}_{\mathrm{SCP}})$ that we obtain from the SDP relaxation $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$.

9.4.1 Cutting plane technique

While the SDP relaxation ($\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I})$) is strongest when we take $\mathcal{I} = \mathcal{I}_{\geq 0}$, it is extremely expensive to keep all the constraints $Y_{ij} \geq 0$. In addition, empirically it is often the case that it suffices to use an appropriate subset $\mathcal{I} \subset \mathcal{I}_{\geq 0}$ whose size is much smaller than $\mathcal{I}_{\geq 0}$ (in the sense that the optimal solution Y of ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) obtained from a SDP solver has zero entries in many $(i,j) \notin \mathcal{I}_{\geq 0}$); but it is often not clear how to pick an "appropriate" subset $\mathcal{I}$ that provide a sufficiently tight SDP relaxation.

To balance the trade-off between the computational costs and using a sufficiently tight SDP relaxation of ($\mathrm{IQP}_{\mathrm{SCP}}$), we employ the following cutting plane technique for finding such an "appropriate" subset $\mathcal{I}$. Intuitively, we put an index (i,j) into $\mathcal{I}$ if it is likely that an optimal solution Y^* of

$$\begin{aligned} \inf_{x,X,Y} \quad & \langle E, X \rangle \\ \text{s.t.} \quad & \left\langle A_{p,\lambda}, Y \right\rangle = 0, \\ & \mathrm{diag}(X) - x = 0, \\ & (A^\top A - I) \circ X = 0, \\ & Y = \begin{bmatrix} 1 & x^\top \\ x & X \end{bmatrix} \succeq 0 \end{aligned} \tag{9.31}$$

has a negative entry at (i,j) (i.e., $Y^*_{ij} < 0$).

Due to the nature of the side chain positioning problem, the matrix E often has a few entries that are much larger in magnitude; intuitively, an optimal solution of (9.31) tends to have a negative entry at (i,j) if $E_{ij} >> 0$ because of the objective (which is to minimize the sum of element-wise products of $\begin{bmatrix} 0 & 0 \\ 0 & E \end{bmatrix}$ and Y). We start with a small initial set $\mathcal{I} \subset \mathcal{I}_{\geq 0}$, where the indices in $\mathcal{I}$ correspond to the largest entries in $\mathcal{E}$. Specifically, in our implementation, we set $\mathcal{I} = \left\{(i,j) \in \mathcal{I}_{\geq 0} : E_{ij} \geq 10^4\right\}$.

Using the initial set $\mathcal{I}$, we solve ($\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I})$) for an optimal solution X^* and take $Y^* = W\hat{X}^*W^\top$. Then we find the indices for the *most violated constraints*. Specifically, we find $(i,j) \in \mathcal{I}_{\geq 0} \backslash \mathcal{I}$ such that $Y^*_{ij} < 0$ and the value $E_{i-1,j-1}Y^*_{ij}$ is very negative (which happens when $E_{i-1,j-1} >> 0$). We update $\mathcal{I}$ by augmenting these new indices, resulting in a slightly larger index set. Then we resolve ($\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I})$).

We fix the number of *cuts* (i.e., the nonnegativity constraints $(W\hat{X}W^\top)_{ij} \geq 0$) to be added in each iteration. The number of cuts incremented in each step has to be chosen with care: an overly small number leads to slow progress, and an overly large number leads to unnecessary

additional computational costs for finding the final solution to ($\mathrm{IQP}_{\mathrm{SCP}}$). As can be expected, the larger the problem is, the larger the per-iteration increment of cuts is required in order to reach a nearly optimal solution for ($\mathrm{IQP}_{\mathrm{SCP}}$) efficiently in practice. In our numerical experiment in Section 9.5, we take the number of cuts to be roughly $0.1n$.

Algorithm 9.1 outlines the subroutine for adding new cutting plane indices. The parameters `tol` and `numcut` represent the tolerance for $Y_{ij} \geq 0$ and the maximum number of cuts added in each iteration.

Algorithm 9.1: Adding Cutting Planes Subroutine, `ACPS`

Parameters(*numcut*, *tol*);
Input($\mathcal{I}$, $\mathcal{I}_{\geq 0}$, $Y \in \mathbb{S}^n_+$ *satisfying* $\{(i,j) \in \mathcal{I}_{\geq 0} : Y_{ij} < \texttt{tol}\} \neq \emptyset$);
Output($\mathcal{I}$)
$\mathcal{I}_{\mathrm{new}} \leftarrow \mathcal{I}_{\geq 0} \cap \{(i,j) : Y_{ij} < \texttt{tol}\}$;
if $|\mathcal{I}_{\mathrm{new}}| >$ *numcut* **then**
    if $E_{ij}Y_{ij} \geq 0$ *for all* $(i,j) \in \mathcal{I}_{\mathrm{new}}$ **then**
        $\mathcal{I}_{\mathrm{new}} \leftarrow$ the set of indices $(i,j) \in \mathcal{I}_{\mathrm{new}}$ for the *numcut*-most negative Y_{ij};
    else
        $\mathcal{I}_{\mathrm{new}} \leftarrow$ the set of indices $(i,j) \in \mathcal{I}_{\mathrm{new}}$ for the *numcut*-most negative $E_{ij}Y_{ij}$;
    endif
endif
$\mathcal{I} \leftarrow \mathcal{I} \cup \mathcal{I}_{\mathrm{new}}$;

9.4.2 Rounding a feasible solution of the SDP relaxation

From a computed optimal solution Y^* of the SDP relaxation ($\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I})$) of the integer program ($\mathrm{IQP}_{\mathrm{SCP}}$), we can obtain a *fractional* solution $c \in \mathbb{R}^{n_0}$ of ($\mathrm{IQP}_{\mathrm{SCP}}$) (i.e., c satisfies $Ac = \bar{e}$ and $c \in [0,1]^{n_0}$) via one of the following common techniques, as in [25].

- *Perron-Frobenius rounding.* Perron-Frobenius rounding uses the best rank-one approximation of Y^*, which by Eckart-Young Theorem is the largest eigenvalue of Y^* times the outer product of the corresponding unit eigenvector [38].

  Let $u \in \mathbb{R}^n$ be an eigenvector corresponding to the largest eigenvalue of Y^*. If $u \geq 0$, then $(u_2, \ldots, u_n)$ is nonzero and the vector $u' := \frac{p}{u_2+\cdots+u_n}(u_2, \ldots, u_n)$ satisfies the constraints $Au' = \bar{e}$ and $u' \geq 0$. Note that if Y^* is nonnegative, then by Perron-Frobenius theorem

(see, e.g., [82]) u is nonnegative. Empirically, even though not all the entries of Y^* are nonnegative, the vector u that we find in the numerical tests are indeed nonnegative.

- *Projection rounding.* The vector $u'' = \mathrm{diag}(Y^*_{2:n,2:n})$ is used. Observe that by Proposition 9.3.7, the feasibility of Y^* in (9.5) implies that u'' satisfies $Au'' = \bar{e}$ and $u'' \geq 0$.

After obtaining a fractional solution c of ($\mathrm{IQP}_{\mathrm{SCP}}$), it is possible to use a probabilistic method to obtain a feasible (integral) solution of ($\mathrm{IQP}_{\mathrm{SCP}}$). Alternatively, we can compute a nearest integral solution x to c, i.e., the nearest vector to c among all the feasible solutions of ($\mathrm{IQP}_{\mathrm{SCP}}$), by solving a linear program.

Proposition 9.4.1. *[24, Proposition 6] For any $c \in \mathbb{R}^{n_0}$, the integer program*

$$\min_x \|x - c\| \quad \text{s.t.} \quad Ax = \bar{e},\ x \in \{0,1\}^{n_0} \tag{9.32}$$

is equivalent to the linear program

$$\min_x -c^\top x \quad \text{s.t.} \quad Ax = \bar{e},\ x \in [0,1]^{n_0}. \tag{9.33}$$

Proof. Any feasible solution x of (9.32) satisfies $x^\top x = p$, so $\|x - c\|^2 = -2c^\top x + (\|c\|^2 + p)$. Therefore (9.32) is equivalent to

$$\min_x -c^\top x \quad \text{s.t.} \quad Ax = \bar{e},\ x \in \{0,1\}^{n_0}. \tag{9.34}$$

Since the columns of A are drawn from the identity matrix I_p, A is totally unimodular. Hence the linear program (9.33), which has a larger feasible region than (9.34) and is feasible and bounded, has an optimal solution that is feasible for (9.34). Therefore (9.34) and (9.33) are equivalent. □

It is clear from the objective of (9.32) that any optimal solution x^* of (9.32) is a greedy solution, in the sense that for each $k \in 1:p$, $x^*_k = e_{i_k}$, where i_k is an index such that the maximum entry of the subvector $c^{(k)}$ lies in the i_k-th position.

9.4.3 Summary of the algorithm

Now we give an explicit description of the heuristic we use for solving ($\mathrm{IQP}_{\mathrm{SCP}}$), in Algorithm 9.2 on Page 158.

We start with an initial index set $\mathcal{I}$ of the cuts. In each iteration of the algorithm, we perform the following steps:

(1) solve ($\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I})$); (we use SDPT3 in our implementation;)

(2) obtain a feasible solution x of ($\mathrm{IQP}_{\mathrm{SCP}}$) using the rounding techniques outlined in Section 9.4.2;

(3) otherwise add new cut indices to $\mathcal{I}$ (using Algorithm 9.1).

Algorithm 9.2 needs the following parameters:

numcut:	number of cuts added in each iteration;
tol:	tolerance for $Y_{ij} \geq 0$;
	(we take *tol* $= 10^{-8}$ in our implementation,
	same as the default tolerance of linear infeasibility in SDPT3;)
maxiter:	maximum number of cutting plane iterations;
r:	maximum number of times an integral solution of ($\mathrm{IQP}_{\mathrm{SCP}}$) is allowed
	to appear consecutively;
ceil_E:	the ceiling on the values of E; (we take *ceil_E* $= 10^5$.)

Algorithm 9.2 takes the following list of input:

p:	number of residues;
$m \in \mathbb{R}^p$:	vector storing the number of rotamers for each residue;
$E \in \mathbb{S}^{n_0}$:	matrix for the objective of ($\mathrm{IQP}_{\mathrm{SCP}}$);
$\mathcal{I} \subset \mathcal{I}_{\geq 0}$:	initial set of indices for the nonnegativity constraint $Y_{ij} \geq 0$ in ($\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I})$).

In the `SDP subroutine`, Any standard SDP solver can be used; we use SDPT3 [92]. Though not stated explicitly in the `SDP subroutine`, we assume that the optimal solution $(t^*, w^*, \Lambda^*, \eta^*)$ of the dual (9.30) is also given alongside the primal optimal solution $\hat{X}^*$. This is a mild assumption, as many standard SDP solvers solve an SDP and its dual simultaneously. As we see in Section 9.5.1, the dual optimal solution is helpful for measuring the quality of the final integral feasible solutions of ($\mathrm{IQP}_{\mathrm{SCP}}$) output by Algorithm 9.2.

9.5 Numerical experiment on some proteins

In this section, we report some numerical results of using the SDP relaxation ($\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I})$) together with the heuristics mentioned in Sections 9.4.1 and 9.4.2, in comparison to using (9.5) from [25], on the reconstruction of 26 proteins listed on the Protein Data Bank.

We first discuss our choice of metrics for measuring the solution quality, in Section 9.5.1. Then we report the metrics for 26 proteins taken from the Protein Data Bank [11] in Section 9.5.2.

Algorithm 9.2 offers superior performance over the direct use of (9.5), thanks to both facial reduction and cutting plane techniques. As a supplement, we also studied the speedup contributed by each of the two techniques using the performance profile [34].

We omit the conformation analysis of the reconstructed proteins in comparison with the Protein Data Bank, which underlines the biological relevance of the solutions we obtained in Section 9.5.2. Interested readers may refer to [24, Section 6.3].

9.5.1 Measuring the quality of feasible solutions of (IQP)

We discuss the metric we use for measuring the quality of a feasible solution u of (IQP_{SCP}) obtained using the SDP relaxation ($\text{P}^{\text{reg}}_{\text{SCP}}(\mathcal{I})$).

Given an SDP solution $(X^*; t^*, \lambda^*, \Lambda^*, \eta^*)$ feasible for ($\text{P}^{\text{reg}}_{\text{SCP}}(\mathcal{I})$)-($\text{D}^{\text{reg}}_{\text{SCP}}(\mathcal{I})$), and from X^*, and a feasible solution u of (IQP_{SCP}) (obtained via, e.g., one of the rounding techniques mentioned in Section 9.4.2), we call the fraction

$$\frac{u^\top Ex - t^*}{\frac{1}{2}|u^\top Eu + t^*|} \tag{9.36}$$

the *relative difference* between the objective value of $(t^*, \lambda^*, \Lambda^*, \eta^*)$ in ($\text{D}^{\text{reg}}_{\text{SCP}}(\mathcal{I})$) and the objective of u in (IQP_{SCP}).

We will using the relative difference defined in (9.36) as a measure of the quality of the feasible solution u of (IQP_{SCP}). Since strong duality holds for ($\text{P}^{\text{reg}}_{\text{SCP}}(\mathcal{I})$) and ($\text{D}^{\text{reg}}_{\text{SCP}}(\mathcal{I})$), they have the same optimal value in theory. In particular,

$$u^\top Eu \geq d_{\mathcal{I}} \geq \bar{d}^*_{\mathcal{I}} \geq t^*,$$

and the smaller the difference $u^\top Eu - t^*$ is, the closer to optimality u is (in (IQP_{SCP})). Ideally, the quantity $|u^\top Eu - v_{\text{scp}}|/|v_{\text{scp}}|$ is a good measure how close to optimality u is; but we usually do not know v_{scp} or each $d_{\mathcal{I}}$ (since the computed SDP solution is only near optimal in general). The only available lower bound for v_{scp} is the objective value of any feasible solution of the dual ($\text{D}^{\text{reg}}_{\text{SCP}}(\mathcal{I})$).

In Section 9.5.2, we also report the dual optimal value (i.e., value of t^*) that we obtain in each instance (although we omit the measure of feasibility of the SDP solutions to save space). As mentioned, t^* provides a lower bound for v_{scp}.

9.5.2 Numerical results

Tables 9.1, 9.2 and 9.3 report the computational results of Algorithm 9.2 versus the direct use of (9.5) to 26 proteins categorized based on the total number of rotamers. The test data were taken from the Protein Data Bank [11] as well as the rotamer library built by the Dunbrack Laboratory [36], and processed using a Python script that executes in the UCSF Chimera molecular modeling environment [71]. (See [24, Section 6.2.1] for further details on problem data generation.)

The following metrics are reported:

- *runtime*. The runtime require for Algorithm 9.2 (which involves solving multiple SDPs) versus solving (9.5) (which require significantly more time since all the nonnegativity constraints are used).

- *dual SDP optval*, the computed optimal value of $(\mathrm{D}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ or equivalently (9.16).

  The dual optimal value serves as a lower bound for v_{scp}.

- *objval in IQP*, the objective value of the computed feasible solution of $(\mathrm{IQP}_{\mathrm{SCP}})$. (We use the Perron-Frobenius rounding.)

  The closer the objective value of the computed feasible solution of $(\mathrm{IQP}_{\mathrm{SCP}})$ is to the computed optimal value of $(\mathrm{D}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$, the closer to optimality the computed feasible solution $(\mathrm{IQP}_{\mathrm{SCP}})$.

- *relative diff*, the relative difference between the computed optimal value of $(\mathrm{D}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$ (or equivalently (9.16)) and the objective value of the computed feasible solution of $(\mathrm{IQP}_{\mathrm{SCP}})$.

- *relative gap*, the relative duality gap of the computed primal-dual optimal solution $(X^*; t^*, \lambda^*, \Lambda^*, \eta^*)$ of $(\mathrm{P}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$-$(\mathrm{D}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$, defined by

$$\frac{(\langle \tilde{E}, X^* \rangle - t^*) + \sum_{(i,j)\in\mathcal{I}} \eta^*_{ij} (W \hat{X}^* W^\top)_{ij}}{1 + |\langle \tilde{E}, X^* \rangle| + |t^*| + |\sum_{(i,j)\in\mathcal{I}} \eta^*_{ij}| + |\sum_{(i,j)\in\mathcal{I}} (W \hat{X}^* W^\top)_{ij}|}.$$

  The relative duality gap measures how accurately the SDP has been solved. (One would expect that the relative duality gap of the SDP solutions used for rounding to be not too big.)

In the numerical tests, Algorithm 9.2 consistently produces better integral solutions in shorter time than the direct use of (9.5). The integer solutions computed by Algorithm 9.2 is essentially optimal because their objective values in $(\mathrm{IQP}_{\mathrm{SCP}})$ are usually very close to the computer optimal values of $(\mathrm{D}^{\mathrm{reg}}_{\mathrm{SCP}}(\mathcal{I}))$, which serve as lower bounds.

Algorithm 9.2: Side Chain Positioning with Cutting Planes, `SCPCP`

Parameters(***numcut***, ***tol***, ***maxiter***, ***r***, ***ceil_E***);

Input($E \in \mathbb{S}^{n_0}$, m, p, $\mathcal{I}$);

Output($u^{(1)}, u^{(2)}$: *feasible solutions for IQP obtained from Perron-Frobenius and projection roundings*);

Initialization;

$n \leftarrow \sum_i m_i + 1$;

$E \leftarrow \min\{E, \textbf{\textit{ceil\_E}}\}$ (element-wise);

$\mathcal{I}_{\geq 0} \leftarrow \{(i,j) : 1 \leq i < j \leq n_0,\ (i,j) \notin \mathcal{B},\ i,j \text{ integral}\}$;

$\tilde{E} \leftarrow W^T \begin{bmatrix} 0 & 0 \\ 0 & E \end{bmatrix} W$;

First iteration;

`SDP subroutine`

- obtain an optimal solution X^* of the optimization problem

$$\begin{array}{ll} \min\limits_{X} & \langle \hat{E}, X \rangle \\ \text{s.t.} & X_{00} = 1,\ \text{arrow}(X) = 0,\ \text{bdiag}(X) = 0, \\ & (WXW^T)_{ij} \geq 0,\ \forall\, (i,j) \in \mathcal{I}, \\ & X \succeq 0, \end{array} \tag{9.35}$$

- $Y^* \leftarrow WX^*W^T$
- obtain $u^{(1)}$ from Perron-Frobenius rounding, and $u^{(2)}$ from projection rounding

More iterations;

for $\ell \in 1 :$ ***maxiter*** **do**

  if $Y^*_{ij} <$ ***tol*** *for some* i, j **then**

    update $\mathcal{I}$ using `Adding Cutting Planes Subroutine` (Algorithm 9.1);

    run `SDP subroutine`;

    if Y^* *is of rank one, or* $u^{(1)}, u^{(2)}$ *are the same as in the previous* ***r*** *iterates* **then**

      STOP;

    endif

  endif

endfor

Table 9.1: Results on small proteins

Protein	n_0	p	run time (sec)		dual SDP optval		objval in IQP		relative diff		relative gap	
			SCPCP	[25]	SCPCP	[25]	SCPCP	[25]	SCPCP	[25]	SCPCP	[25]
1AAC	117	85	6.58	296.06	-206.33	-206.33	-206.33	-206.33	5.75E-11	1.72E-05	1.30E-09	4.21E-04
1AHO	108	54	7.97	364.73	33.53	33.53	33.53	33.53	8.44E-11	4.95E-05	2.45E-09	4.68E-04
1BRF	130	45	14.96	977.08	-31.11	-31.11	-31.11	-31.11	3.92E-11	2.27E-05	3.08E-09	1.24E-04
1CC7	160	66	28.60	1059.06	-63.76	-2.30E+07	-63.76	3.73E+04	1.13E-11	2.01	1.27E-09	1.11
1CKU	115	60	5.46	815.18	113.83	113.83	113.83	113.83	7.17E-11	4.79E-05	3.42E-09	1.13E-04
1CRN	65	37	12.76	46.42	-14.87	-14.87	-14.87	-14.87	1.64E-12	3.05E-05	2.20E-10	3.66E-04
1CTJ	153	61	16.15	777.31	-129.53	-6.69E+06	-129.53	174.65	2.98E-11	2.00	2.29E-09	1.07
1D4T	188	89	41.32	2775.34	-173.03	-2.96E+07	-173.03	291.13	3.88E-11	2.00	1.35E-09	1.20
1IGD	82	50	5.51	189.04	-69.25	-69.25	-69.25	-69.25	4.79E-10	2.74E-06	5.76E-09	3.39E-05
1PLC	129	82	14.32	1766.03	-1.50	-1.50	-1.50	-1.50	1.28E-11	7.28E-04	4.60E-10	1.09E-03
1VFY	134	63	23.49	1765.36	-90.09	-90.09	-90.09	-90.09	1.67E-11	-1.11E-05	9.15E-10	3.79E-05
4RXN	98	48	18.44	366.48	-21.65	-21.65	-21.65	-21.65	1.48E-11	2.62E-05	4.19E-10	6.67E-05

Table 9.2: Results on medium-sized proteins

Protein	n_0	p	run time (min)		dual SDP optval		objval in IQP		relative diff		relative gap	
			SCPCP	[25]	SCPCP	[25]	SCPCP	[25]	SCPCP	[25]	SCPCP	[25]
1B9O	265	112	0.64	254.85	-140.24	-5.63E+07	-140.24	1.91E+06	1.19E-11	2.14	1.45E-09	1.24
1C5E	200	71	2.59	70.63	-131.75	-6.46E+04	-131.75	148.82	4.93E-11	2.01	5.02E-09	1.00
1C9O	207	53	2.15	66.50	-83.55	-1.88E+06	-83.55	1628.10	3.35E-12	2.00	2.77E-10	1.02
1CZP	237	83	1.90	143.95	-37.88	-2.26E+04	-37.88	1254.42	8.30E-11	2.24	1.03E-08	1.00
1MFM	216	118	0.19	102.11	-201.29	-7.36E+07	-201.29	1369.92	2.01E-11	2.00	1.24E-09	1.09
1QQ4	365	143	5.70	-	-102.40	-	-102.40	-	6.49E-11	-	2.27E-08	-
1QTN	302	134	5.04	-	-178.77	-	-178.77	-	2.24E-11	-	4.12E-09	-
1QU9	287	101	7.55	-	-124.96	-	-124.96	-	1.80E-11	-	5.52E-09	-

Table 9.3: Results on large proteins (SCPCP only)

Protein	n_0	p	run time (hr)	dual SDP optval	Objval in IQP	rel. diff	rel. gap	numcut	# iter	Final # cuts
1CEX	435	146	0.08	140.20	140.20	1.26E-11	5.57E-09	40	9	485
1CZ9	615	111	3.96	497.46	497.46	2.98E-13	6.37E-10	60	25	1997
1QJ4	545	221	0.15	-286.83	-286.83	5.31E-12	1.14E-09	60	14	1027
1RCF	581	142	0.85	-191.54	-191.54	3.71E-12	1.15E-08	60	17	1305
2PTH	930	151	29.65	-159.41	-159.41	8.69E-09	7.63E-06	120	34	7247
5P21	464	144	0.31	-135.75	-135.75	1.39E-12	7.33E-10	40	16	822

9.5.3 Individual speedup contributed by facial reduction and cutting planes

One reason why the solution of (9.5) requires considerably longer time is the formidable amount of nonnegativity constraints. As a supplement, we study the speedup contributed by each of the two techniques, the facial reduction and the cutting plane, using the performance profile [34]. Specifically, we consider the four different methods:

(1) SCPCP, i.e., Algorithm 9.2 (the facial reduction and the cutting plane techniques combined),

(2) only the cutting plane technique,

(3) only the facial reduction, and

(4) the original SDP relaxation (9.5).

For each of the 26 instance, i.e., $i \in 1:26$ and each method $j \in 1:4$, define

$$\begin{aligned} t_{i,j} &:= \text{ run time for getting the final solution of IQP for instance } i \text{ by method } j, \\ r_{i,j} &:= \frac{t_{i,j}}{\min\{t_{i,j} : j = 1,2,3,4\}}. \end{aligned}$$

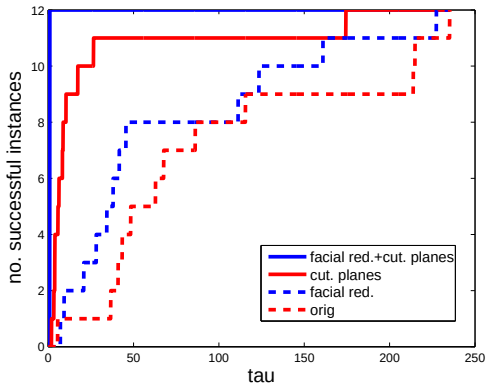
The fraction $r_{i,j}$ is called the *performance ratio* of method j on instance i, and measures how much worse the run time of method j is over the best method on the same instance. The ratio is at least 1; the larger the ratio is, the worse the method performs relative to the best method.

The performance profile in Figure 9.1 plots (on the vertical axis) the number ρ_j defined by

$$\rho_j(\tau) := \text{ number of instance } i \text{ such that } r_{i,j} \le \tau$$

for each method j, for $j = 1, \ldots, 4$, as $\tau \ge 0$ increases, among all the small protein (listed in Table 9.1).

Figure 9.1: Performance profile comparing the four methods



As expected, Algorithm 9.2 has the best run time. While the performance profile indicates that facial reduction plays a smaller role in the speedup than the cutting plane technique does, the importance of the facial reduction in regularizing the original SDP formulation ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) should not be underestimated, especially when the size of the problem instance (i.e., n) is large, in which case the numerical instability of ($\mathrm{P}_{\mathrm{SCP}}(\mathcal{I})$) becomes very prominent and it is impossible to get any reasonable solution.

Chapter 10

Conclusion

This thesis studies the use of facial reduction in regularizing semidefinite programs that are not strictly feasible. We considered an implementation of the facial reduction algorithm for semidefinite programs and some associated numerical issues; we showed that each iteration of the facial reduction algorithm is backward stable. Then we gave an overview of some uses of the facial reduction in very different areas, from the theoretical results such as error bounds for linear matrix inequalities and sensitivity analysis of semidefinite programming, to applications such as sensor network localization and reducing the size of SDP relaxations from different discrete problems. In particular, we studied the use of facial reduction in regularizing a SDP relaxation of the NP-hard side chain positioning problem from protein folding.

Facial reduction highlights the importance of good modeling, i.e., that one should try to make use of as much information available as possible before writing down the optimization problem. In both theory (when dealing with, e.g., SDP relaxations from certain problems) and practice (when solving an SDP numerically using a solver) it is important to check whether strict feasibility holds for the SDP as well as its dual. Without such a safety check, an unsuspicious user may end up using excessive amounts of computation time, only to obtain a solution that may be far from feasible, as we saw in Sections 5.5 and 9.5.2. Often times, the failure of strict feasibility suggests that the underlying model may not adequately capture all the mathematical features of the problem. As mentioned in Section 8.1.1, in some occasions the failure of strict feasibility could have been avoided if the model is more refined (in the sense that more valid constraints are used). From this perspective, the use of facial reduction is indeed of a mending nature, to improve the given model and make explicit certain hidden features.

10.1 Future directions

An interesting future direction would be to use the facial reduction to understand the "complexity" of some geometric objects. Sturm used the facial reduction algorithm to provide an error bound result for linear matrix inequalities; can it be extended to other more general convex sets? In the proof, the linearity of the inequalities played a rather important role (specifically, allowing for the use of Hoffman's error bound). The knowledge of the facial structure was also very important; one can possibly extend the result to, for instance, feasible regions of second order cone programs, as the second order cone has an even simpler facial structure.

Another subject not fully addressed in this thesis is the facial structure associated with conic program over a Cartesian products of cones. In the simplest case, if $\mathcal{K}$ in the conic program ($\mathrm{P_m}$) is given by $\mathbb{R}^{n_1}_+ \times \mathbb{R}^{n_2}_+ \times \cdots \mathbb{R}^{n_k}_+$, then it is clear that ($\mathrm{P_m}$), being a linear program, requires at most one iteration of facial reduction. What if $\mathcal{K} = \mathbb{S}^{n_1} \times \cdots \times \mathbb{S}^{n_k}$ for example? Would ($\mathrm{P_m}$) also require at most one iteration of facial reduction, as ($\mathrm{P_{SOCP}}$) does (as in Theorem 4.2.1)? More generally, if each inequality $C^{(j)} - (\mathcal{A}^{(j)})^* y \in \mathcal{K}_j$ in ($\mathrm{P_m}$) is strictly feasible by itself, can the failure of strict feasibility be removed after one iteration of facial reduction? Since the facial structure of a Cartesian product of cones is no more complicated than the facial structure of the constituent cones (in Prop 2.2.19), it seems that the answer should be positive.

Finally, we remark that there is more that can be done in the area of polynomial optimization. The facial reduction algorithm considered in [94] does not exactly correspond to the graph theoretical approach in [58], which constructs a graph and removes one vertex at a time, and essentially is a special case of the facial reduction, where in each iteration the dimension of the smaller face found goes down by exactly one. It is also interesting to note that, in the case of sparse SOS representation, there is a correspondence between the facial reduction and a graph "reduction", that has not been fully explored.

References

[1] T. Akutsu. NP-hardness results for protein side-chain packing. *Genome Informatics*, 8:180–186, 1997. 129, 132

[2] B. Alipanahi, N. Krislock, A. Ghodsi, H. Wolkowicz, L. Donaldson, and M. Li. Determining protein structures from NOESY distance constraints by semidefinite programming. *J. Comput. Biol.*, 20(4):296–310, 2013. 3

[3] F. Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM J. Optim.*, 5(1):13–51, 1995. 1

[4] F. Alizadeh and D. Goldfarb. Second-order cone programming. *Math. Program.*, 95(1, Ser. B):3–51, 2003. ISMP 2000, Part 3 (Atlanta, GA). 30

[5] M.F. Anjos and J.B. Lasserre. Introduction to semidefinite, conic and polynomial optimization. In *Handbook on semidefinite, conic and polynomial optimization*, volume 166 of *Internat. Ser. Oper. Res. Management Sci.*, pages 1–22. Springer, New York, 2012. 1, 27, 31

[6] G.P. Barker. The lattice of faces of a finite dimensional cone. *Linear Algebra and Appl.*, 7:71–82, 1973. 17

[7] G.P. Barker. Faces and duality in convex cones. *Linear and Multilinear Algebra*, 6(3):161–169, 1978/79. 17

[8] G.P. Barker. Theory of cones. *Linear Algebra Appl.*, 39:263–291, 1981. 17

[9] G.P. Barker and D. Carlson. Cones of diagonally dominant matrices. *Pacific J. Math.*, 57(1):15–32, 1975. 17

[10] S.J. Benson and Y. Ye. Algorithm 875: DSDP5—software for semidefinite programming. *ACM Trans. Math. Software*, 34(3):Art. 16, 20, 2008. 1

[11] H.M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T.N. Bhat, H. Weissig, I.N. Shindyalov, and P.E. Bourne. The protein data bank. *Nucleic Acids Res*, 28:235–242, 2000. 156, 157

[12] G. Blekherman, P.A. Parrilo, and R.R. Thomas, editors. *Semidefinite Optimization and Convex Algebraic Geometry*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 2012. 1, 15, 27, 32

[13] V. Boltyanski, H. Martini, and P.S. Soltan. *Excursions into combinatorial geometry*. Universitext. Springer-Verlag, Berlin, 1997. 17

[14] J.F. Bonnans and A. Shapiro. Optimization problems with perturbations: a guided tour. *SIAM Rev.*, 40(2):228–264, 1998. 94

[15] B. Borchers. CSDP 2.3 user's guide. *Optim. Methods Softw.*, 11/12(1-4):597–611, 1999. Interior point methods. 1

[16] B. Borchers. CSDP, a C library for semidefinite programming. *Optim. Methods Softw.*, 11/12(1-4):613–623, 1999. Interior point methods. 1

[17] J.M. Borwein and A.S. Lewis. *Convex analysis and nonlinear optimization*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 3. Springer, New York, second edition, 2006. Theory and examples. 99

[18] J.M. Borwein and H. Wolkowicz. Characterization of optimality for the abstract convex program with finite-dimensional range. *J. Austral. Math. Soc. Ser. A*, 30(4):390–411, 1980/81. 2

[19] J.M. Borwein and H. Wolkowicz. Facial reduction for a cone-convex programming problem. *J. Austral. Math. Soc. Ser. A*, 30(3):369–380, 1980/81. 2

[20] J.M. Borwein and H. Wolkowicz. Regularizing the abstract convex program. *J. Math. Anal. Appl.*, 83(2):495–530, 1981. 2, 41, 47

[21] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear matrix inequalities in system and control theory*, volume 15 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. 112, 125

[22] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004. 1, 13, 27

[23] R.E. Burkard, E. Çela, P. M. Pardalos, and L.S. Pitsoulis. The quadratic assignment problem. In *Handbook of combinatorial optimization, Vol. 3*, pages 241–237. Kluwer Acad. Publ., Boston, MA, 1998. 113

[24] F. Burkowski, Y.-L. Cheung, and H. Wolkowicz. Efficient use of semidefinite programming for selection of rotamers in protein conformations. Submitted to INFORMS J. Comput., 2013. 3, 112, 131, 142, 144, 146, 148, 150, 154, 156, 157

[25] B. Chazelle, C. Kingsford, and M. Singh. A semidefinite programming approach to side chain positioning with new rounding strategies. *INFORMS J. Comput.*, 16(4):380–392, 2004. 112, 121, 129, 131, 132, 135, 141, 150, 153, 155, 159

[26] J.T.W. Cheng and S. Zhang. On implementation of a self-dual embedding method for convex programming. *Optim. Methods Softw.*, 21(1):75–103, 2006. 2

[27] Y-L. Cheung, S. Schurr, and H. Wolkowicz. Preprocessing and regularization for degenerate semidefinite programs. In D.H. Bailey, H.H. Bauschke, P. Borwein, F. Garvan, M. Thera, J. Vanderwerff, and H. Wolkowicz, editors, *Computational and Analytical Mathematics, In Honor of Jonathan Borwein's 60th Birthday*, volume 50 of *Springer Proceedings in Mathematics & Statistics*. Springer, 2013. x, 2, 3, 47, 60, 61, 67, 70, 79, 81, 85

[28] M.D. Choi, T.Y. Lam, and B. Reznick. Sums of squares of real polynomials. In *K-theory and algebraic geometry: connections with quadratic forms and division algebras (Santa Barbara, CA, 1992)*, volume 58 of *Proc. Sympos. Pure Math.*, pages 103–126. Amer. Math. Soc., Providence, RI, 1995. 122

[29] E. de Klerk, D.V. Pasechnik, and R. Sotirov. On semidefinite programming relaxations of the traveling salesman problem. *SIAM J. Optim.*, 19(4):1559–1573, 2008. 112, 117

[30] E. de Klerk, C. Roos, and T. Terlaky. Initialization in semidefinite programming via a self-dual skew-symmetric embedding. *Oper. Res. Lett.*, 20(5):213–221, 1997. 2, 4

[31] E. de Klerk, C. Roos, and T. Terlaky. Infeasible-start semidefinite programming algorithms via self-dual embeddings. In *Topics in semidefinite and interior-point methods (Toronto, ON, 1996)*, volume 18 of *Fields Inst. Commun.*, pages 215–236. Amer. Math. Soc., Providence, RI, 1998. 2, 4

[32] J. Demmel and B. Kågström. The generalized Schur decomposition of an arbitrary pencil $A - \lambda B$: robust software with error bounds and applications. II. Software and applications. *ACM Trans. Math. Software*, 19(2):175–201, 1993. 77

[33] P.J.C. Dickinson. *The Copositive Cone, the Completely Positive Cone and their Generalisations*. PhD thesis, University of Groningen, 2013. 14, 31

[34] E.D. Dolan and J.J. Moré. Benchmarking optimization software with performance profiles. *Math. Program.*, 91(2, Ser. A):201–213, 2002. 156, 160

[35] R.J. Duffin. Infinite programs. In *Linear inequalities and related systems*, Annals of Mathematics Studies, no. 38, pages 157–170. Princeton University Press, Princeton, N. J., 1956. 95

[36] R.L. Dunbrack, Jr. and M. Karplus. Backbone-dependent rotamer library for proteins application to side-chain prediction. *Journal of Molecular Biology*, 230(2):543–574, March 1993. 157

[37] M. Dür, B. Jargalsaikhan, and G. Still. The Slater condition is generic in linear conic programming, 2012. 2

[38] C. Eckart and G. Young. The approximation of one matrix by another of lower rank. *Psychometrika*, 1:211–218, 1936. 153

[39] R.M. Freund. Complexity of an algorithm for finding an approximate solution of a semidefinite program with no regularity assumption. Technical Report OR-302-94, Operations Research Center, M.I.T., Cambridge, MA, USA, December 1994. 1

[40] R.M. Freund, R. Roundy, and M.J. Todd. Identifying the set of always-active constraints in a system of linear inequalities by a single linear program. Technical Report WP 1674-85, MIT Sloan School of Management, Cambridge, MA, USA, October 1985. 62

[41] R.M. Freund and J.R. Vera. Some characterizations and properties of the "distance to ill-posedness" and the condition measure of a conic linear system. *Math. Program.*, 86(2, Ser. A):225–260, 1999. 36

[42] K. Fujisawa, M. Kojima, and K. Nakata. The interior-point method software SDPA (semidefinite programming algorithm) for semidefinite programming problems. *Sūrikaisekikenkyūsho Kōkyūroku*, pages 149–159, 1999. Continuous and discrete mathematics for optimization (Kyoto, 1999). 1

[43] K. Fujisawa, K. Nakata, M. Yamashita, and M. Fukuda. SDPA project: solving large-scale semidefinite programs. *J. Oper. Res. Soc. Japan*, 50(4):278–298, 2007. 1

[44] B. Ghaddar. *A Branch-and-Cut Algorithm based on Semidefinite Programming for the Minimum k-Partition Problem*. PhD thesis, University of Waterloo, 2007. 129, 130, 132

[45] D. Goldfarb and K. Scheinberg. On parametric semidefinite programming. *Appl. Numer. Math.*, 29(3):361–377, 1999. Proceedings of the Stieltjes Workshop on High Performance Optimization Techniques (HPOPT '96) (Delft). 94

[46] A.J. Goldman and A.W. Tucker. Theory of linear programming. In *Linear inequalities and related systems*, Annals of Mathematics Studies, no. 38, pages 53–97. Princeton University Press, Princeton, N.J., 1956. 36

[47] G.H. Golub and C.F. Van Loan. *Matrix computations*. Johns Hopkins Studies in the Mathematical Sciences. Johns Hopkins University Press, Baltimore, MD, fourth edition, 2013. 75

[48] M. Grant and S. Boyd. Graph implementations for nonsmooth convex programs. In V. Blondel, S. Boyd, and H. Kimura, editors, *Recent Advances in Learning and Control*, Lecture Notes in Control and Information Sciences, pages 95–110. Springer-Verlag Limited, 2008. `http://stanford.edu/~boyd/graph_dcp.html`. 127

[49] G. Gruber and F. Rendl. Computational experience with ill-posed problems in semidefinite programming. *Comput. Optim. Appl.*, 21(2):201–212, 2002. 3

[50] B. Grünbaum. *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Günter M. Ziegler. 17

[51] B. Hendrickson. *The Molecule Problem: Determining Conformation from Pairwise Distances*. PhD thesis, Cornell University, 1991. 124

[52] B. Hendrickson. Conditions for unique graph realizations. *SIAM J. Comput.*, 21(1):65–84, 1992. 124

[53] N.J. Higham. *Accuracy and stability of numerical algorithms*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 2002. 82, 125

[54] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex analysis and minimization algorithms. I*, volume 305 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1993. Fundamentals. 8, 15, 17

[55] R.A. Horn and C.R. Johnson. *Topics in matrix analysis*. Cambridge University Press, Cambridge, 1994. Corrected reprint of the 1991 original. 125

[56] CVX Research Inc. CVX: Matlab software for disciplined convex programming, version 2.0. `http://cvxr.com/cvx`, August 2012. 127

[57] B. Jansen, C. Roos, and T. Terlaky. The theory of linear programming: skew symmetric self-dual problems and the central path. *Optimization*, 29(3):225–233, 1994. 1

[58] M. Kojima, S. Kim, and H. Waki. Sparsity in sums of squares of polynomials. *Math. Program.*, 103(1, Ser. A):45–62, 2005. 112, 122, 123, 163

[59] N. Krislock. *Semidefinite Facial Reduction for Low-Rank Euclidean Distance Matrix Completion*. PhD thesis, University of Waterloo, 2010. 112, 124

[60] N. Krislock and H. Wolkowicz. Explicit sensor network localization using semidefinite representations and facial reductions. *SIAM J. Optim.*, 20(5):2679–2708, 2010. 3, 112, 124

[61] A.S. Lewis. Facial reduction in partially finite convex programming. *Math. Programming*, 65(2, Ser. A):123–138, 1994. 2

[62] Z.-Q. Luo, J.F. Sturm, and S. Zhang. Duality results for conic convex programming. Report 9719/A, Econometric Institute, Erasmus University Rotterdam, 1997. 2, 17, 32, 44, 95

[63] Z.-Q. Luo, J.F. Sturm, and S. Zhang. Conic convex programming and self-dual embedding. *Optim. Methods Softw.*, 14(3):169–218, 2000. 2, 4

[64] A.M. Lyapunov. *The general problem of the stability of motion*. Taylor & Francis Ltd., London, 1992. 125

[65] Yu. Nesterov and A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*, volume 13 of *SIAM Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1994. 1

[66] Yu. Nesterov, M. J. Todd, and Y. Ye. Infeasible-start primal-dual methods and infeasibility detectors for nonlinear programming problems. *Math. Program.*, 84(2, Ser. A):227–267, 1999. 2

[67] P.A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Math. Program.*, 96(2, Ser. B):293–320, 2003. Algebraic and geometric methods in discrete optimization. 122

[68] G. Pataki. On the closedness of the linear image of a closed convex cone. *Math. Oper. Res.*, 32(2):395–412, 2007. 32

[69] G. Pataki. Bad semidefinite programs: they all look the same, 2011. 97

[70] G. Pataki. Strong duality in conic linear programming: facial reduction and extended duals. *ArXiv e-prints*, January 2013. 2, 47

[71] E.F. Pettersen, T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin. UCSF Chimera–a visualization system for exploratory research and analysis. *Journal of computational chemistry*, 25(13):1605–1612, oct 2004. 157

[72] V. Powers and T. Wörmann. An algorithm for sums of squares of real polynomials. *J. Pure Appl. Algebra*, 127(1):99–104, 1998. 122

[73] M.V. Ramana. An exact duality theory for semidefinite programming and its complexity implications. *Math. Programming*, 77(2, Ser. B):129–162, 1997. Semidefinite programming. 2, 96

[74] M.V. Ramana, L. Tunçel, and H. Wolkowicz. Strong duality for semidefinite programming. *SIAM J. Optim.*, 7(3):641–662, 1997. 2

[75] J. Renegar. Some perturbation theory for linear programming. *Math. Programming*, 65(1, Ser. A):73–91, 1994. 36

[76] James Renegar. Incorporating condition measures into the complexity theory of linear programming. *SIAM J. Optim.*, 5(3):506–524, 1995. 36

[77] Bruce Reznick. Extremal PSD forms with few terms. *Duke Math. J.*, 45(2):363–374, 1978. 122

[78] R.T. Rockafellar. *Conjugate duality and optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1974. 94

[79] R.T. Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks. 8, 11, 15, 17

[80] J.B. Saxe. Embeddability of weighted graphis in k-space is strongly NP-hard. In *Proceedings of the 17th Allerton Conference on Communications, Control, and Computing*, pages 480–489. University of Illinois at Urbana-Champaign, 1979. 124

[81] A. Schrijver. *Theory of linear and integer programming.* Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons Ltd., Chichester, 1986. A Wiley-Interscience Publication. 17

[82] E. Seneta. *Non-negative matrices and Markov chains.* Springer Series in Statistics. Springer, New York, 2006. Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544]. 154

[83] A. Shapiro. On duality theory of convex semi-infinite programming. *Optimization*, 54(6):535–543, 2005. 99

[84] J.F. Sturm. *Primal dual interior point approach to semidefinite programming.* PhD thesis, Erasmus University Rotterdam, 1997. 44

[85] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optim. Methods Softw.*, 11/12(1-4):625–653, 1999. Interior point methods. 1, 2, 3, 79

[86] J.F. Sturm. Error bounds for linear matrix inequalities. *SIAM J. Optim.*, 10(4):1228–1248 (electronic), 2000. 2, 5, 6, 17, 32, 47, 93, 101, 105

[87] M. Tanaka, K. Nakata, and H. Waki. Application of a facial reduction algorithm and an inexact primal-dual path-following method for doubly nonnegative relaxation for mixed binary nonconvex quadratic optimization problems. *Pac. J. Optim.*, 8(4):699–724, 2012. 2

[88] M.J. Todd. Semidefinite optimization. *Acta Numer.*, 10:515–560, 2001. 1, 27, 40

[89] L. Tunçel. On the Slater condition for the SDP relaxations of nonconvex sets. *Oper. Res. Lett.*, 29(4):181–186, 2001. 116, 138

[90] L. Tunçel. *Polyhedral and semidefinite programming methods in combinatorial optimization*, volume 27 of *Fields Institute Monographs.* American Mathematical Society, Providence, RI, 2010. 1, 13, 27, 58, 95, 115, 116

[91] L. Tunçel and H. Wolkowicz. Strong duality and minimal representations for cone optimization. *Comput. Optim. Appl.*, 53(2):619–648, 2012. ix, 19, 20, 32, 35, 65, 79, 80

[92] R.H. Tütüncü, K.C. Toh, and M.J. Todd. Solving semidefinite-quadratic-linear programs using SDPT3. *Math. Program.*, 95(2, Ser. B):189–217, 2003. Computational semidefinite and second order cone programming: the state of the art. 1, 3, 127, 155

[93] R.J. Vanderbei. *Linear programming*. International Series in Operations Research & Management Science, 114. Springer, New York, third edition, 2008. Foundations and extensions. 29

[94] H. Waki and M. Muramatsu. A facial reduction algorithm for finding sparse SOS representations. *Oper. Res. Lett.*, 38(5):361–365, 2010. 3, 122, 123, 163

[95] H. Waki and M. Muramatsu. Facial reduction algorithms for conic optimization problems. *J. Optim. Theory Appl.*, 158(1):188–215, 2013. 2, 47

[96] H. Waki, M. Nakata, and M. Muramatsu. Strange behaviors of interior-point methods for solving semidefinite programming problems in polynomial optimization, 2008. 3

[97] H. Wei and H. Wolkowicz. Generating and measuring instances of hard semidefinite programs. *Math. Program.*, 125(1, Ser. A):31–45, 2010. ix, 79, 80

[98] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of semidefinite programming*. International Series in Operations Research & Management Science, 27. Kluwer Academic Publishers, Boston, MA, 2000. Theory, algorithms, and applications. 1, 13, 27

[99] H. Wolkowicz and Q. Zhao. Semidefinite programming relaxations for the graph partitioning problem. *Discrete Appl. Math.*, 96/97:461–479, 1999. The satisfiability problem (Certosa di Pontignano, 1996); Boolean functions. 3

[100] M. Yamashita, K. Fujisawa, M. Fukuda, K. Kobayashi, K. Nakata, and M. Nakata. Latest developments in the SDPA family for solving large-scale SDPs. In *Handbook on semidefinite, conic and polynomial optimization*, volume 166 of *Internat. Ser. Oper. Res. Management Sci.*, pages 687–713. Springer, New York, 2012. 1

[101] Y. Ye. *Interior point algorithms*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons Inc., New York, 1997. Theory and analysis, A Wiley-Interscience Publication. 1, 2, 4

[102] Y. Ye, M.J. Todd, and S. Mizuno. An $O(\sqrt{n}L)$-iteration homogeneous and self-dual linear programming algorithm. *Math. Oper. Res.*, 19(1):53–67, 1994. 1

[103] F. Zhang, editor. *The Schur complement and its applications*, volume 4 of *Numerical Methods and Algorithms*. Springer-Verlag, New York, 2005. 13

[104] S. Zhang. A new self-dual embedding method for convex programming. *J. Global Optim.*, 29(4):479–496, 2004. 2, 4

[105] Q. Zhao, S.E. Karisch, F. Rendl, and H. Wolkowicz. Semidefinite programming relaxations for the quadratic assignment problem. *J. Comb. Optim.*, 2(1):71–109, 1998. Semidefinite programming and interior-point approaches for combinatorial optimization problems (Toronto, ON, 1996). 3, 112, 113, 115

Index